STRUCTURAL BASIS OF VIRAL RNA RECOGNITION BY RIG-I-LIKE RECEPTORS

A Dissertation

by

CHENG LU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Biochemistry

Structural Basis of Viral RNA Recognition by RIG-I-Like Receptors

STRUCTURAL BASIS OF VIRAL RNA RECOGNITION

BY RIG-I-LIKE RECEPTORS

A Dissertation

by

CHENG LU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Pingwei Li
Committee Members,	David P. Barondeau
	Tatyana Igumenova
	Xiuren Zhang
Head of Department,	Gregory D. Reinhart

August 2012

Major Subject: Biochemistry

ABSTRACT

Structural Basis of Viral RNA Recognition by RIG-I-Like Receptors. (August 2012)

Cheng Lu, B.S., Nankai University, China

Chair of Advisory Committee: Dr. Pingwei Li

RIG-I-like receptors (RLRs), RIG-I, MDA5, and LGP2, are a family of innate immune receptors that recognize viral RNA in the cytoplasm and initiate antiviral responses including the induction of type I interferons and other pro-inflammatory cytokines. All three proteins have both an RNA helicase domain with ATPase activity and a C-terminal domain (CTD) which is responsible for RNA binding. RIG-I and MDA5 also have two tandem caspase activation and recruitment domains (CARDs) at the N-terminus which are involved in downstream signaling.

To understand the structural basis of viral RNA recognition by the RLRs, especially RIG-I, we have performed extensive biochemical studies to determine the binding properties of RIG-I with different forms of RNA, including dsRNA with and without 5'-triphosphate (5'-ppp) groups, and 5'-ppp ssRNA. RIG-I CTD binds to these forms of RNA, and exhibits the highest affinity for 5'-ppp dsRNA. We also determined the crystal structures of RIG-I CTD in complex with dsRNA with and without 5'-ppp by X-ray crystallography. The structures showed that RIG-I recognizes the termini of the dsRNA and interacts with the two types of RNA in different orientations. By comparing these complex structures together with mutagenesis studies, we conclude that RIG-I

CTD is a versatile binding module capable of recognizing different RNA ligands. Similar but partially differing sets of residues are involved in the recognition of dsRNA with and without 5'-ppp. Mutations of key residues at the RNA binding surface also abolished RIG-I signaling in cells.

In order to compare the RIG-I/RNA interactions with other RLRs, we also determined the dsRNA binding surface of MDA5 CTD by NMR titration studies. MDA5 CTD has a similar binding surface to that of RIG-I CTD, however with slightly different surface electrostatic potentials which indicate different interactions with RNA. This may explain how MDA5 senses differing types of viruses compared to RIG-I.

The current RIG-I activation model suggests that after stimulation by RNA binding, RIG-I undergoes an ATP-dependent conformational change, exposing the CARDs for downstream signaling. To understand the critical role that the helicase domain plays in RIG-I activation by structural approach, we also attempted to crystallize the dsRNA-bound helicase domain together with CTD.

DEDICATION

To my parents and my husband, for their endless support and love.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Pingwei Li, for his guidance and support throughout these years. I am also thankful to all my other committee members: Dr. Xiuren Zhang, Dr. Tatyana Igumenova, Dr. David Barondeau, and Dr. Sumana Datta, for their helpful suggestions and encouragement.

I convey my thanks to my current and previous lab mates, especially Dr. Xiaojun Li, Dr. Lujiang Hao, and Dr. Chang Shu, for their constant support and precious help in my work. Thanks also go to our wonderful collaborators, Dr. Cheng Kao, Dr. Andrew Herr, and Dr. Roland Strong, for their efforts in this project.

I also want to extend my gratitude to all my friends, colleagues, and the department faculty and staff, for making my time at Texas A&M University a great experience.

Thank you.

NOMENCLATURE

A	Alanine
ADAR1	Adenosine deaminase acting on RNA-1
ADP	Adenosine-5'-diphosphate
AIDS	Acquired immune deficiency syndrome
AMP-PNP	Adenosine 5'-(β,γ-imido)triphosphate
APOBEC3G	Apolipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like 3G
Arg	Arginine
Asp	Aspartate
ASU	Asymmetric unit
ATF	Activating transcription factor
ATP	Adenosine-5'-triphosphate
ATPase	Adenosine triphosphatase
AUC	Analytical ultracentrifugation
BIR	Baculovirus inhibitor of apoptosis repeat
C	Cysteine
CARD	Caspase activation and recruitment domain
Cardif	CARD adaptor inducing interferon-β
cDNA	Complementary deoxyribonucleic acid
CpG	Cytosine-phosphate-guanine

CTD	Carboxyl-terminal domain
Cys	Cysteine
D	Aspartate
DC	Dendritic cell
DEN	Dengue virus
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
dsDNA	Double-stranded deoxyribonucleic acid
dsRNA	Double-stranded ribonucleic acid
E	Glutamate
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EGCG	Epigallocatechin-3-gallate
elF2α	E74-like factor 2α
EMCV	Encephalomyocarditis virus
EMSA	Electrophoretic mobility shift assay
EtBr	Ethidium bromide
F	Phenylalanine
FADD	Fas-associated protein with death domain
Glu	Glutamate
H	Histidine
HD	Helicase domain

HEK	Human embryonic kidney
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HSQC	Heteronuclear single quantum correlation
IFN	Interferon
IFNR	Interferon receptor
IKK	Inhibitor of κB kinase
IL	Interleukin
Ile	Isoleucine
IL-1R	Interleukin-1 receptor
IPS-1	Interferon promoter stimulator 1
IPP	Inorganic pyrophosphatase
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
IRF	Interferon regulatory factor
ISRE	Interferon-stimulated response element
JEV	Japanese encephalitis virus
K	Lysine
LB	Luria Bertani broth
Leu	Leucine
LGP2	Laboratory of genetics and physiology 2
LPS	Lipopolysaccharide
LRR	Leucine rich repeat

Lys	Lysine
mAb	Monoclonal antibody
MAVS	Mitochondrial antiviral signaling protein
MDA5	Melanoma differentiation-associated gene 5
mRNA	Messenger ribonucleic acid
MyD88	Myeloid differentiation primary response gene 88
NCS	Non-crystallography symmetry
NDV	Newcastle disease virus
NF-κB	Nuclear factor κB
NK	Natural killer cell
NLR	Nucleotide-oligomerization-domain-like receptor
NMR	Nuclear magnetic resonance
OAS	2'-5'-Oligoadenylate synthetase
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen-associated molecular pattern
PCR	Polymerase chain reaction
PDB	Protein data bank
pDC	Plasmacytoid dendritic cell
PEG	Polyethylene glycol
Phe	Phenylalanine
PKR	Protein kinase R
PMSF	Phenylmethylsulfonyl fluoride

Poly I:C	Polyinosinic:polycytidylic acid
pp	Diphosphate
ppp	Triphosphate
PRR	Pattern recognition receptor
PYD	Pyrin domain
R	Arginine
RIG-I	Retinoic acid-inducible gene-I
RLR	RIG-I-like receptor
r.m.s.d.	Root mean square deviation
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
RNase	Ribonuclease
rNTP	Ribonucleoside triphosphate
rRNA	Ribosomal ribonucleic acid
S	Serine
SAP	Shrimp alkaline phosphatase
SARS	Severe acute respiratory syndrome
Sc	Shape correlation statistics
SDS	Sodium dodecyl sulfate
Ser	Serine
SeV	Sendai virus
SF2	Superfamily 2

Sf9	Spodoptera frugiperda 9
SPR	Surface plasmon resonance
ssRNA	Single-stranded ribonucleic acid
TBK1	TANK-binding kinase 1
TIR	Toll/IL-1R
TIRAP	TIR-domain-containing adaptor protein
TLR	Toll-like Receptor
TRADD	Tumor necrosis factor receptor type 1-associated death domain protein
TRAF3	Tumor necrosis factor receptor-associated factor 3
TRIF	TIR-domain-containing adapter inducing interferon-β
Tris	2-Amino-2-hydroxymethyl-1,3-propanediol
tRNA	Transfer ribonucleic acid
Trp	Tryptophan
Tyr	Tyrosine
VISA	Virus-induced signaling adapter
VSV	Vesicular stomatitis virus
X	Any amino acid residue

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

The innate immune system and pattern recognition receptors

Both animals and plants are constantly threatened by the invasion of microorganisms including bacteria, viruses, fungi, and parasites, which cause many kinds of infectious diseases. Therefore immune defense systems have been developed to combat and eliminate infective pathogens (1,2). The mammalian immune system effectively fights against pathogens through the cooperation of two branches, innate and adaptive (acquired) immunities. The innate immunity provides the first line of host defense against pathogens at the early stage of infection, with contributions from macrophages, dendritic cells (DCs), mast cells, neutrophils, eosinophils, and natural killer (NK) cells, etc. (3). Adaptive immunity is involved in the elimination of pathogens in later phases of infection as well as the generation of lasting immunological memory. It is characterized by specific recognition which is achieved by B and T lymphocytes bearing an indefinite number of antigen-specific receptors generated by gene rearrangement. The adaptive immune responses are relatively slow processes that often require days or weeks to reach optimal effects. During this period, the innate immune response plays a critical role in controlling infections. It also initiates and subsequently directs adaptive immune responses (4).

This dissertation follows the style of *Nucleic Acids Research*.

Although the innate immune system lacks the fine specificity of adaptive immunity that is necessary to produce immunological memory, it is not completely nonspecific. It is able to discriminate between noninfectious "self" and infectious "non-self". This is achieved by utilizing a limited number of germline-encoded pattern recognition receptors (PRRs) that are constitutively expressed in all cells of a given type, either on the cell surface, in intracellular compartments, or secreted into the bloodstream and tissue fluids (5). They recognize specific microbial components which are highly conserved and are essential for the survival or pathogenicity of the microorganism, such as flagellin, sugars, or the cell wall components peptidoglycan and lipopolysaccharide (LPS). These are known as pathogen-associated molecular patterns (PAMPs) because they are shared by large groups of pathogens but do not exist in the host organism. The binding of PAMPs by PRRs results in very rapid responses, while the adaptive immune response requires a delay for the clonal expansion of lymphocytes. Once activated, the PRRs initiate the pro-inflammatory signaling pathways, leading to the secretion of humoral factors known as cytokines and chemokines, inducing the expression of adhesion and costimulatory molecules in order to recruit immune cells to the site of infection and to trigger the adaptive immune response (6).

Several types of PRRs have been identified. They can be classified as membrane-bound PRRs, cytoplasmic PRRs, and secreted PRRs. For example, macrophage mannose receptor and macrophage scavenger receptor are cell surface receptors functioning as PRRs (7,8), while serum amyloid protein, mannan-binding lectin, and C-reactive protein

are secreted PRRs produced by the liver during the acute phase response at the early stage of infection (7,9).

The most intensely studied and probably the best characterized PRRs are the mammalian Toll-like receptors (TLRs) that are homologues of Drosophila Toll. These are a family of membrane-bound receptors localizing either to the cell surface or to endosomal compartments. TLRs are type I integral membrane glycoproteins characterized by the extracellular domains containing multiple leucine-rich-repeat (LRR) motifs that form a horseshoe fold, and a cytoplasmic signaling domain homologous to that of the interleukin-1 receptor (IL-1R), termed the Toll/IL-1R homology (TIR) domain (10). Ten different TLRs have been identified in humans and twelve in mice (11-13), detecting various ligands including lipids (TLR1, TLR2, and TLR6), flagellin (TLR5), double-stranded (ds) viral RNA (TLR3), single-stranded (ss) viral RNA (TLR7/TLR8), and bacterial DNA (TLR9). Current models assume that TLRs are activated through a ligand-induced dimerization that brings the cytosolic TIR domains in close proximity, which enables the recruitment of adaptor proteins, such as MyD88, TRIF, and TIRAP, for downstream signaling (14).

Some pathogens can gain access to the intracellular compartments, such as the cytosol. There are also some cytosolic PRRs, e.g. the nucleotide oligomerization domain (NOD)-like receptors (NLRs) and RIG-I-like receptors (RLRs), which can detect these intracellular pathogens and induce responses that block their replication. The NLRs are multidomain proteins with a tripartite architecture containing C-terminal LRRs for ligand sensing, a central nucleotide-binding NACHT domain whose oligomerization is

necessary for protein activation, and an N-terminal effector domain which can be a caspase activation and recruitment domain (CARD), pyrin domain (PYD), or baculovirus IAP (inhibitor of apoptosis) repeat (BIR) domain (6). PYDs and CARDs are structurally related and belong to the death-fold domain superfamily whose members are frequently found in pathways leading to the activation of caspases or the nuclear factor κB (NF-κB) (15). The death-fold domains are characterized by six α helices that are tightly packed in a Greek key fold and form trimers or dimers with other members of the same subfamily (e.g. either PYD/PYD or CARD/CARD interactions) to mediate signal transduction to downstream targets.

Critical roles of type I interferons in innate immunity toward viral infection

Viral pathogens have been discovered in all species from bacteria to mammals. Viruses cause many kinds of infectious diseases such as the common cold, influenza, measles, rabies, hepatitis, many forms of diarrhea, yellow fever, polio, smallpox, SARS, and AIDS. A number of viral diseases such as polio and smallpox have almost been eliminated by successful immunization in early childhood. However, many other viral diseases such as hepatitis and AIDS are still major threats to human health, especially for people in the developing world.

Viral infections provoke various host responses, including early innate and subsequent adaptive immune responses. For innate responses, some PRRs are genetically programmed to detect a wide range of viral infections and activate the induction of cytokines and chemokines. The most important cytokines in viral infection

are type I interferons (IFNs), including a single IFN-β and multiple subtypes of the IFN-α family(16), that are secreted and activate target cells in autocrine and paracrine manners, conferring antiviral activity to the host, by inducing apoptosis in virus-infected cells, and rendering cells resistant to the infection (17-19). Furthermore, IFN and other cytokines critically contribute to the successful activation of the adaptive immunity, and the stimulation of hematopoietic stem cell turnover and proliferation (20).

The immune responses mediated by IFNs are regulated by two signaling cascades: a primary IFN producing signal induced by pathogens and a secondary signal mediated by IFN receptors (21,22). The genes encoding IFNs are regulated by the coordinate and cooperative assembly of an enhanceosome containing several transcription factors including activating transcription factor (ATF)/c-*jun*, NF-κB, and IFN regulatory factors (IRFs), which are activated by virus-induced signaling (23-26). The IFNs expressed in response to viral infections relay a secondary signal through interaction with the IFN receptors (IFNRs) on the cell surface and induce the expression of hundreds of IFN-inducible genes, such as genes encoding dsRNA-dependent protein kinase R (PKR), 2'-5'-oligoadenylate synthetase (OAS), RNA-specific adenosine deaminase ADAR1 (adenosine deaminase acting on RNA-1), and RNA-editing enzyme APOBEC3G (27). PKR phosphorylates the translation initiation factor elF2α, leading to the down regulation of mRNA translation (28). The 2'-5' OAS activates ribonuclease RNase L, promoting viral RNA degradation and stimulating antiviral responses (29,30). ADAR-1 and APOBEC3G deaminate viral RNA, leading to mutations of the viral genome (31). Together, these IFN-inducible proteins down-regulate protein synthesis,

promote cell growth arrest and induce apoptosis, thus creating an "antiviral state" to contain the spread of viruses (18). Furthermore, type I IFNs enhance the maturation of DCs and activate NK cells as well as macrophages. IFNs also stimulate antibody production by B cells and trigger the expansion and differentiation of cytotoxic T cells, facilitating the adaptive immune responses (32-34).

Viral nucleic acids as common PAMPs recognized by PRRs during viral infections

Viruses contain genetic material composed of either DNA or RNA that encodes viral structural components as well as synthetic and replication enzymes. Besides surface glycoproteins, these viral nucleic acids are commonly recognized as PAMPs by the innate immune system because of their specific structures that differ from the host DNA or RNA. For example, most viral DNA genomes contain unmethylated cytosine-guanine (CpG) dinucleotide sequences at a high frequency, while mammalian DNA has a low frequency of CpG dinucleotides and most are methylated. In addition, most viruses produce unique dsRNA, and some viruses produce uracil-rich ssRNA during portions of their life cycle.

Several proteins that are involved in the recognition of viral nucleotides have been discovered, including the TLR family members TLR3, TLR7/8, and TLR9 described above. TLR3 detects viral dsRNA and triggers TRIF-mediated pathways, leading to IRF-3 and NF-κB activation and IFN-β induction (35). Viral ssRNA and DNA are recognized by TLR7/8 and TLR9, respectively, and stimulate MyD88-mediated pathways, resulting in IRF-7 activation and IFN-α induction (36).

Because the TLRs are membrane-bound proteins, they can only recognize extracellular dsRNA or ssRNA associated with viral particles that are internalized into the endosomes by endocytosis. Nevertheless, in most types of cells the accumulation of intracellular dsRNA resulting from the replication of viruses also triggers host response mechanisms including the expression of type I IFN, through a TLR3-independent pathway (37). Later, three DExD/H box RNA helicases were identified as a family of cytosolic receptors sensing intracellular dsRNA. The prototypical member retinoic acid-inducible gene-I (RIG-I), together with melanoma differentiation-associated gene 5 (MDA5; also called Helicard) and laboratory of genetics and physiology 2 (LGP2), were named as RIG-I like receptors (RLRs). The RLR family becomes the central of research in these few years, as it is essential in the recognition of cytosolic viral RNA and the induction of type I IFNs and other pro-inflammatory cytokines such as interleukin-6 (IL-6), conferring antiviral activity to the host cells and activating the adaptive immune responses.

Given the critical roles of type I IFNs in innate immunity against viruses, the detection of viruses and the stimulation of IFNs secretion by RLRs and TLRs are very important for the antiviral innate immune responses (18,38).

RIG-I and MDA5: virus sensing and signaling

RNA viruses comprise approximately 80% of all viruses and in humans are etiologic agents of infectious diseases causing major public health concerns (39). RLRs and some TLRs are involved in the recognition of viral RNA. Whereas TLRs detect viral

components in specific cell types, such as macrophages and DCs, RLRs sense the infected viruses in the cytoplasm of most cell types except plasmacytoid DCs (pDCs) (40). In contrast to the TLRs described above, RLRs lack repetitive receptor elements such as LRRs which could serve as a recognition platform. Therefore, the finding that RLRs were involved in viral sensing in the cytosol was unexpected (41).

The RLRs are DExD/H-box RNA helicases containing DECH motif and belong to the helicase superfamily 2 (SF2). The DExD/H box RNA helicases, defined by their ability to unwind dsRNA with their intrinsic ATPase activity, are found in almost all organisms from viruses to mammals and are involved in a wide variety of biological functions such as transcription, translation, pre-mRNA splicing, and RNAi (42-44).

RIG-I was first isolated as a gene induced by retinoic acid in a promyelocytic leukemia cell line (45). It was then identified as a cDNA clone that activates IRF-regulated reporter gene expression in response to transfection of the synthetic dsRNA analogue polyinosinic-polycytidylic acid (poly I:C) (41). Depletion of RIG-I inhibits IFN-β production in response to Newcastle disease virus (NDV), suggesting that RIG-I is required for virus-triggered signaling.

MDA5, as the closest relative of RIG-I, was originally identified in human melanoma cells as a gene activated by IFN-β and protein kinase C-activating compound mezerein (46). MDA5 was then linked to the mediation of antiviral responses and identified as a binding target for V proteins of paramyxoviruses (47,48).

Both RIG-I and MDA5 consist of two tandem CARDs at the N-terminus, a centrally located DExD/H-box RNA helicase domain with ATPase activity, and a C-

terminal repressor or regulatory domain (CTD, or RD) (Figure 1.1) (49,50). RIG-I and MDA5 share 23% and 44% amino acid identities in the N-terminal tandem CARDs and the helicase domain, respectively. The CARDs of RIG-I and MDA5 participate in the signaling cascade. Overexpression of the CARDs alone is sufficient to induce antiviral signaling resulting in IFN production. The CTDs and the helicase domains are involved in viral RNA sensing and the regulation of RLR activation (49,51). The helicase domain retains catalytic activity to unwind dsRNA in an ATP hydrolysis-dependent manner. However, the unwinding activity may be reversibly correlated with the antiviral signaling activity (52).

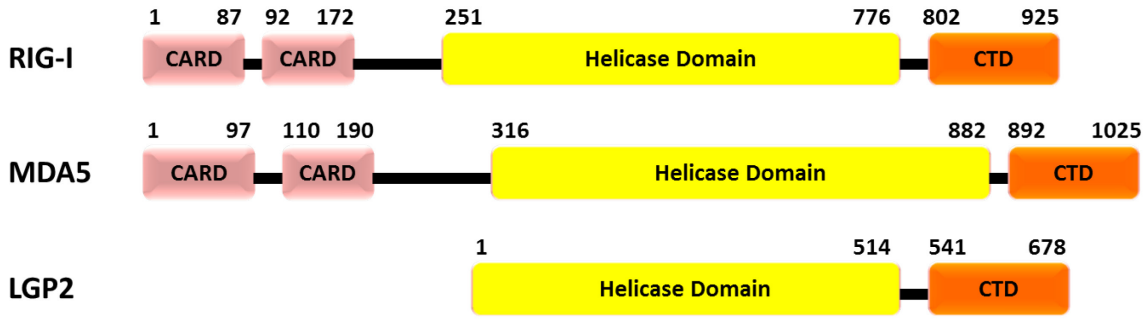


Figure 1.1 Domain structures of the RIG-I-like receptors.
RIG-I and MDA5 are composed of two tandem CARDs at the N-terminus, a central DECH helicase domain, and a C-terminal domain. Lacking the N-terminal CARDs, LGP2 shows a homologous domain architecture with only a helicase domain and a CTD.

Studies of RIG-I and MDA5 knockout mice demonstrated the differential roles of these two receptors in the recognition of RNA viruses (53). RIG-I is essential for the induction of IFNs in response of a set of RNA viruses including the paramyxoviruses such as NDV and Sendai virus (SeV), influenza A virus, vesicular stomatitis virus (VSV), and Japanese encephalitis virus (JEV) (40). In contrast, MDA5 is critical for the detection of picornaviruses, such as encephalomyocarditis virus (EMCV), Mengo virus, and Theiler's virus (54). Some viruses such as Dengue virus (DEN), West Nile virus and reovirus are recognized by both receptors redundantly (55). Since the induction of type I IFNs by various RNA viruses is completely abolished in cells lacking both RIG-I and MDA5, the RLRs are likely sole cytosolic sensors of RNA viruses regulating type I IFN induction (53).

LGP2: a regulator of RIG-I and MDA5 signaling

The third RLR family member LGP2 was originally cloned as a gene neighboring the Stat3 and Stat5 loci on mouse chromosome 11, which encodes a cytoplasmic protein of previously unknown function (56). Later it was found to be a homolog of RIG-I and MDA5, with 31% and 41% amino acid identities to the helicase domains of RIG-I and MDA5, respectively, but it completely lacks CARDs (Figure 1.1). Therefore it has no direct signaling capability and functions as a regulator of RIG-I and MDA5 signaling (57,58). LGP2 has low basal expression level at steady state, but is strongly inducible by antiviral stimulations such as poly I:C transfection, virus infection, and IFN-α treatment. LGP2 exhibits high affinity for both dsRNA and 5'-ppp ssRNA and

serves as a negative regulator of RIG-I signaling (57,59), either by sequestering dsRNA, inhibiting RIG-I conformational changes, or through a signaling pathway. On the other hand, LGP2 seems to enhance responses to poly I:C mediated by MDA5 (60). The controversial roles of LGP2 in antiviral signaling need to be explained in future studies.

RNA ligands of RLRs

Being innate immune receptors, the RLRs are able to discriminate between the self (host) and non-self (viral) RNAs. In addition, RIG-I and MDA5 respond to distinct but overlapping sets of viruses, suggesting that they specifically recognize different forms of virus RNA. In order to elucidate the specific RNA recognition by each of the RLRs, extensive studies have been carried out and several types of RNA ligands have been identified.

For example, 5'-ppp ssRNA is recognized exclusively by RIG-I. Two initial studies using RNA transcribed from DNA templates by T7 RNA polymerase demonstrated that RNA containing 5'-ppp groups potently triggers the activation of RIG-I upon transfection into various types of cells, and it was suggested that 5'-ppp ssRNA is the primary ligand of RIG-I (61,62). While the genomes of most RNA viruses contain 5'-ppp RNA, the 5'-ppp moieties of most host RNA are either masked by a 7-methyl-guanosine cap (in the case of mRNA) or removed during RNA processing (tRNA and rRNA) (63). Therefore the self RNA can avoid triggering the RLRs signaling cascade.

However, recent research has revealed that 5'-ppp *per se* may not be sufficient for RIG-I activation and a dsRNA structure may be required. Studies using chemically

synthesized dsRNA or hairpin RNA containing 5'-ppp demonstrated that dsRNA with 5'-ppp is a strong activator of RIG-I and the base-pairing near the 5'-ppp is necessary (64,65). On the other hand, other studies showed that dsRNA without 5'-ppp can also activate RIG-I signaling in transfected cells (66). In addition, the panhandle structure of the influenza virus genome was also recently reported as a RIG-I ligand (67). Because different forms of RNA, cell lines, and assays were used in these studies, questions remain about the exact structural features of viral RNA recognized by RIG-I.

The synthetic dsRNA poly I:C was known as a potent non-viral IFN inducer and interacts with TLR3 (68). Poly I:C is generated by annealing poly I and poly C, and harbors a 5'-diphosphate (5'-pp) (69). Commercial poly I:C migrates as dsRNA of 4-8 kb in gel electrophoresis. Analysis of RIG-I and MDA5 knockout mice demonstrated that MDA5, but not RIG-I, is responsible for IFN induction in response to poly I:C (53,54,70). Interestingly, partial digestion of poly I:C with RNase III, which generates trimmed poly I:C of about 300 bp, transforms it from a ligand for MDA5 to a ligand for RIG-I, suggesting that MDA5 recognizes long dsRNA while RIG-I prefers short dsRNA (71). Consistent with this speculation, long dsRNA (> 2 kb) purified from EMCV infected cells can trigger MDA5-dependent IFN production. The reoviral genome consists of 10 segments with different sizes, including long (3.9 kb), medium (2.2-2.3 kb), and short (1.2-1.4 kb). The long and short segments are ligands for MDA-5 and RIG-I, respectively (71). It is not clear how RIG-I and MDA5 discriminate between short and long dsRNA at the molecular level.

The structural basis of viral RNA recognition by RLRs

Extensive protease treatment of the RIG-I/dsRNA complex followed by mass spectrometry and protein sequencing analysis identified that the CTD of RIG-I (residues 792-925) is involved in dsRNA and 5'-ppp ssRNA recognition (52). The affinity of RIG-I CTD for dsRNA or 5'-ppp ssRNA is slightly lower than that of full-length RIG-I, indicating that the helicase domain may also play some role in RNA binding (52,72). However, the helicase domain of RIG-I alone does not bind to RNA (52).

To understand the structural basis of viral RNA recognition by RLRs, the structures of RIG-I, MDA5 and LGP2 CTDs have been determined by X-ray crystallography and nuclear magnetic resonance (NMR) (52,60,72,73). Although the amino acid sequence of RIG-I CTD is about 30% and 29% identical to the CTDs of MDA5 and LGP2 respectively (Figure 1.2), the structures of the three RLR CTDs are very similar and can be superimposed with only slight differences in the flexible loop regions (Figure 1.3). The CTD is a flat domain with a convex and a concave side and dimensions of about 45 Å × 35 Å × 30 Å. Four conserved cysteine residues coordinate a zinc atom which plays a crucial role in maintaining proper folding and functioning of the CTD. The structures revealed a highly conserved RNA binding cleft in the CTD of RLRs (49). NMR titrations of RIG-I CTD indicated that a large positively charged surface located at the center of the structure is involved in both 5'-ppp ssRNA and dsRNA recognition (52). Mutagenesis of several positively charged residues on this surface either reduced or disrupted RNA binding by RIG-I and LGP2 (52,60,72).

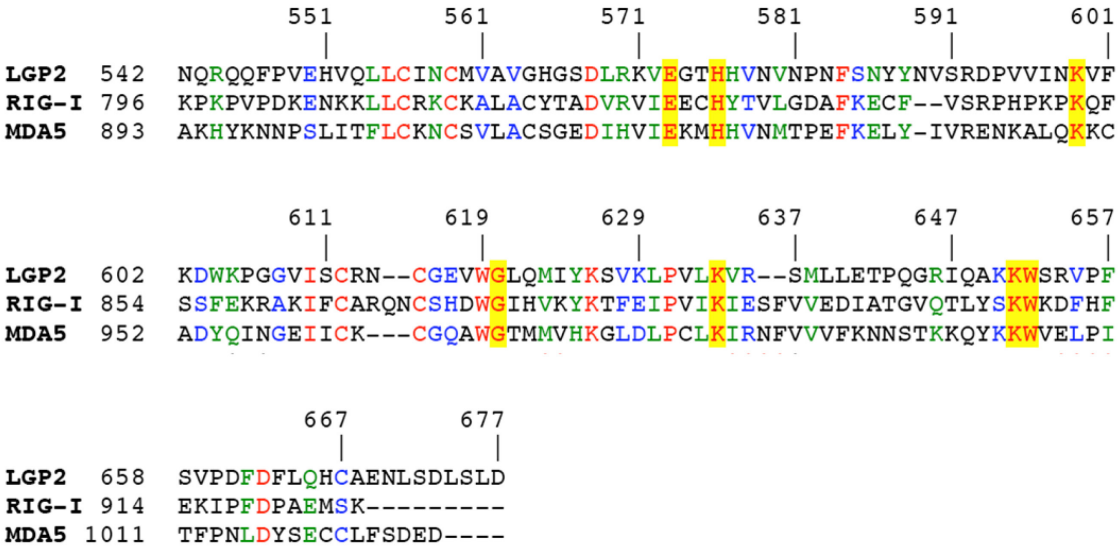


Figure 1.2 Sequence alignment of human RIG-I, MDA5, and LGP2 CTDs. Residues conserved in all three proteins are colored red. Conserved residues that are involved in dsRNA binding surface are highlighted in yellow. Conservatively replaced residues are in green and blue.

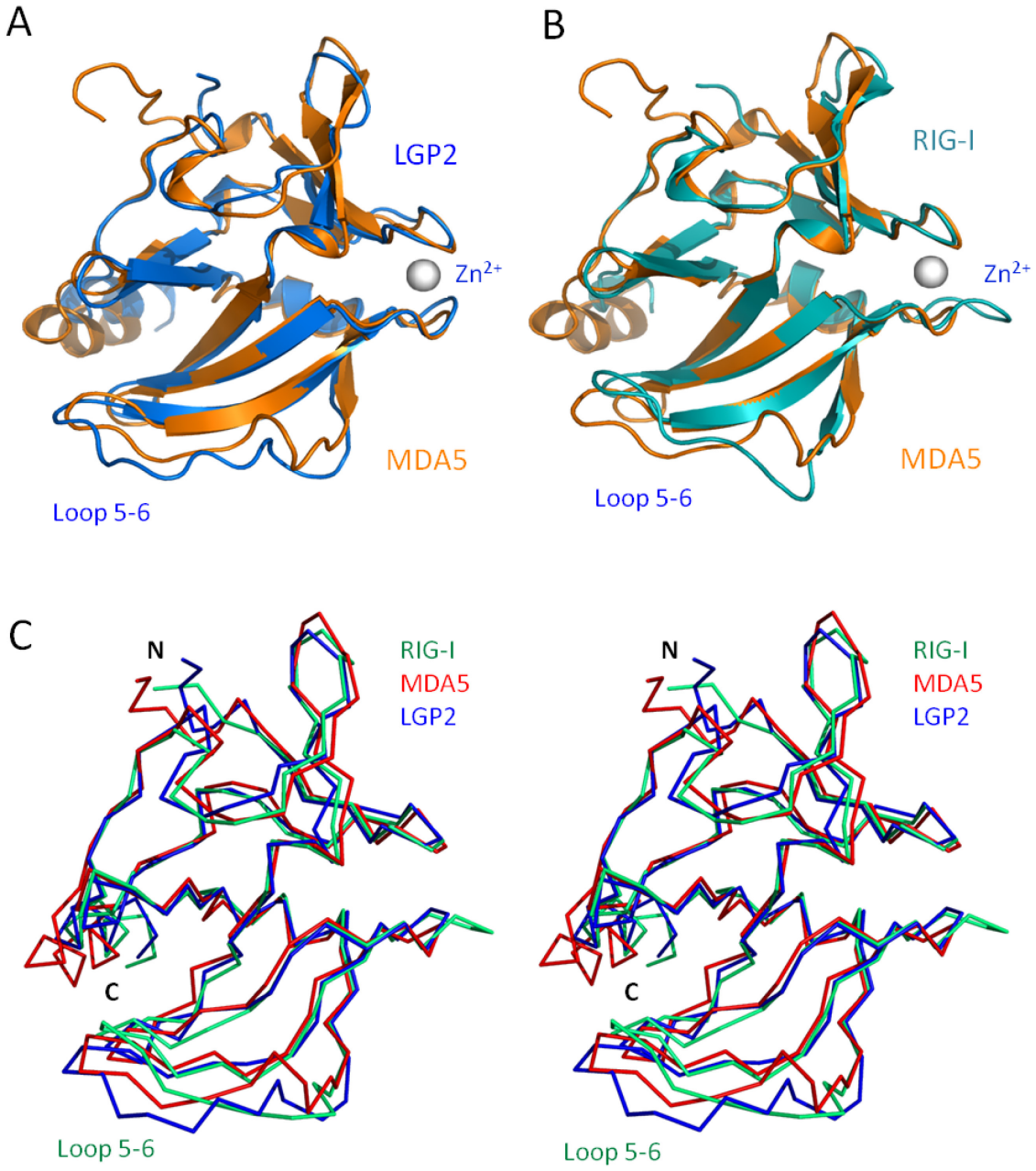


Figure 1.3 Comparison of CTD structures of RLRs.

(A) Ribbon model of superimposed MDA5 and LGP2 CTDs.

(B) Ribbon model of superimposed RIG-I and MDA5 CTDs.

(C) Stereo view of superimposed α-Carbon of three RLR CTDs.

The crystal structure of LGP2 CTD bound to an 8-bp dsRNA, which was the first CTD/RNA complex structure, was later determined in our laboratory (74), revealing that LGP2 specifically recognizes the blunt ends of dsRNA. This provides the first insight into how the RLRs interact with dsRNA. The recognition of the terminal structures of viral RNA allows the receptors to discriminate signature structures of viral RNA and detect the level of infection by sensing the concentration of viral RNA as well. This mechanism of viral RNA sensing by the RLRs suggests a way to fine-tune the immune response to viral infection and avoid damage to the host cells. Consistent with findings from the LGP2/dsRNA complex structure, luciferase reporter gene assays showed that RIG-I was activated by short dsRNA with blunt ends but not by dsRNA with either 3'- or 5'-overhangs (74).

Mechanism of RLR activation, signaling, and regulation

Overexpression of full-length RIG-I exhibits low levels of basal activity (58). Furthermore, although IFN treatment strongly induces RIG-I protein, IFN treatment alone does not activate IFN genes. Thus, an autorepression model has been proposed (Figure 1.4). Without RNA activation, full-length RIG-I is unable to induce downstream signaling, possibly due to intramolecular inhibition by the CTD, which mediates a closed structure (38,49). The engagement of viral RNA by RIG-I induces an ATP-dependent conformational change, exposing the CARDs for the recruitment of downstream adaptor protein interferon promoter stimulator 1 (IPS-1, also known as MAVS, VISA, or Cardif), which is located on the outer membrane of mitochondria and contains an N-

terminal CARD (75-78). Activated RIG-I interacts with IPS-1 through CARD-CARD interaction (38,51), serving as a platform for the assembly of signaling complexes involving TRAF3 and TRADD (38,79). TRAF3 leads to the activation of the noncanonical IκB (inhibitor of κB) kinase-ε (IKKε) and TANK-binding kinase-1 (TBK1) (80) that activate IRF3 and IRF7 by phosphorylation (81,82). Activated IRF3 and IRF7 translocate to the nucleus and bind to the IFN-stimulated response elements (ISREs), inducing the transcription of type I IFN genes (26,80,83). The TRADD and FADD arm of the signaling pathway are involved in the activation of transcription factors NF-κB and IRFs that regulate the expression of both pro-inflammatory cytokines and type I IFNs (84,85). In addition, post-translational modifications of RIG-I, including ubiquitination, phosphorylation, and SUMOylation, have been reported to be important for the activation or regulation of RIG-I signaling (86,87). Moreover, RIG-I dimerization or oligomerization is also likely to be critical for its function (88).

The autoregulation by the CTD does not appear to be conserved in MDA5, since the expression of MDA5 in cells results in high basal signaling activity in the absence of RNA, although the addition of RNA ligand increases MDA5 signaling (89).

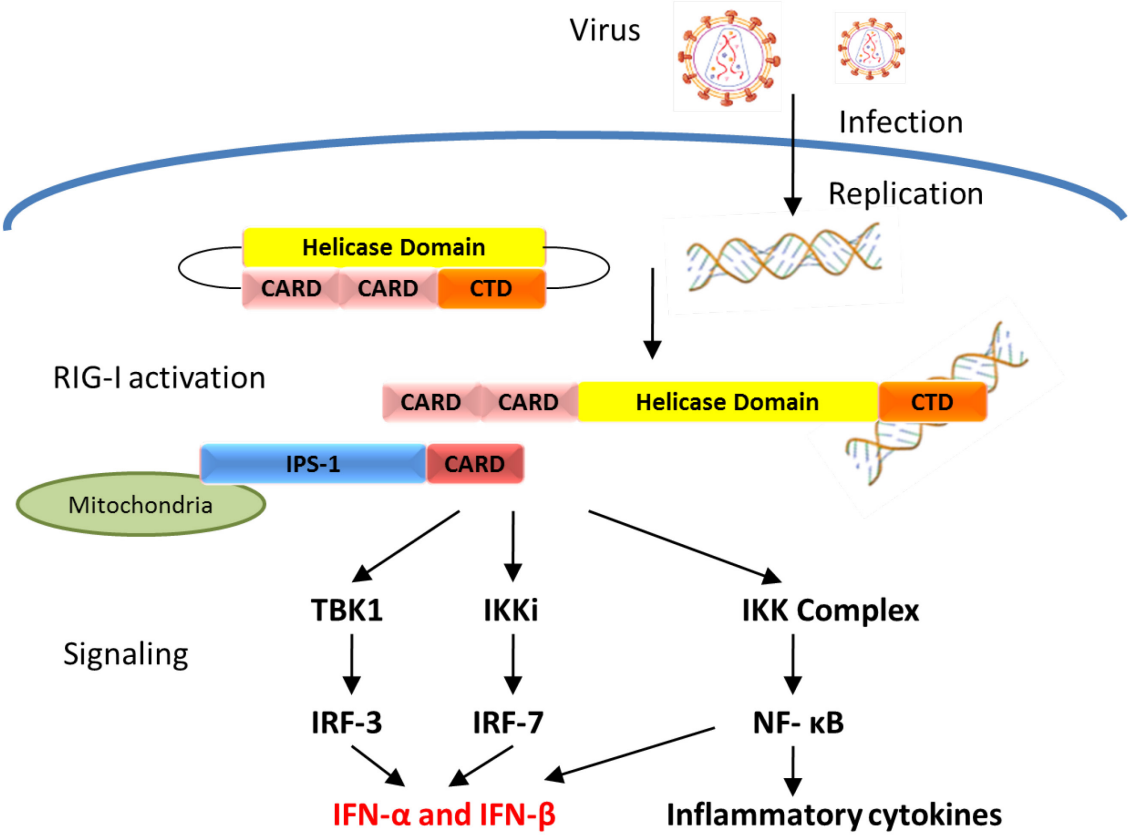


Figure 1.4 Model of RIG-I activation and signaling.

RIG-I is expressed in cytoplasm as an inactive form, in which CARDs are masked sterically by intra-molecular interaction with CTD. After viral infection, viral RNA binds to RIG-I and induces a conformational change in the presence of ATP. CARDs are exposed and interact with a downstream adaptor, IPS-1 (also called MAVS, VISA and Cardif), on mitochondria. IPS-1 interacts with various adaptor molecules leading to the activation of regulatory kinases, TBK1, IKKi (ε), and the IKKα/β/γ complex. These kinases activate respective transcription factors, IRF-3, IRF-7, and NF-κB, which participate in the activation of IFN and other cytokine genes.

Overview of dissertation chapters

Our general interest is to investigate the structural basis of the interactions between RLR receptors and their RNA ligands, and the activation of RLR antiviral signaling. With Chapter I as an introduction to the innate immune receptors RLRs, in the following chapters the structural and biochemical studies of two RLR proteins, RIG-I and MDA5, are described. This dissertation is primarily focused on the prototypical RLR protein RIG-I in Chapters II, III, and V, with an additional study performed on MDA5 in Chapter IV.

Chapter II and III are focused on the interactions between RIG-I CTD and various RNA ligands. Examinations of the RNA ligand specificities and binding properties of RIG-I have been performed by extensive biochemical methods. High resolution crystal structures of RIG-I CTD bound to dsRNA with 5'-ppp group are described in Chapter II, while the crystal structure of RIG-I CTD bound to non-ppp dsRNA is described in Chapter III. The structural studies together with mutagenesis studies elucidate the detailed interactions between RIG-I CTD and different RNA ligands, and a comparison of them is discussed as well.

Chapter IV describes the RNA binding properties of MDA5. Residues of MDA5 CTD that are involved in dsRNA binding were mapped by an NMR titration study, showing a similar but also distinctive RNA-binding surface compared to RIG-I CTD. This enables the comparison of different binding interactions between RLRs and RNA ligands.

Chapter V discusses the cloning, protein overexpression, purification, crystallization trials, and preliminary X-ray diffraction data of a longer version of RIG-I consisting of its helicase domain and the CTD. This would provide insights into the role of RIG-I helicase domain plays in RNA detection and RIG-I activation for signaling.

Chapter VI is the final chapter providing a brief summary on the research in this dissertation as well as the future directions that can still be pursued.

CHAPTER II

STRUCTURAL BASIS OF 5'-TRIPHOSPHATE DOUBLE-STRANDED RNA RECOGNITION BY RIG-I C-TERMINAL DOMAIN*

Background

RIG-I is an innate immune receptor which detects viral RNA from replicating viruses in the cytoplasm and induces the secretion of type I IFNs (IFN-α and –β) and other pro-inflammatory cytokines, conferring antiviral activity to the host cells and activating acquired immune responses (18,19,34,38). RIG-I is responsible for sensing NDV, SeV, influenza A virus, VSV, and JEV, etc. (40). However, the types of RNA ligands that are specifically recognized by RIG-I are still under debate, and the mechanism of viral RNA recognition is not fully understood.

The 5'-ppp group was initially identified as a key structural feature of viral RNA that was sensed by RIG-I, and it was suggested that 5'-ppp ssRNA is the primary ligand of RIG-I (61,62,90). However, recent studies demonstrated that dsRNA with 5'-ppp are potent activators of RIG-I (64,65). In addition, the crystal structure of human LGP2 CTD bound to dsRNA showed that the termini of dsRNA are recognized by LGP2 (74).

---

*Reproduced with permission from "The structural basis of 5' triphosphate double-stranded RNA recognition by RIG-I C-terminal domain" by Lu C., Xu H., Ranjith-Kumar C.T., Brooks M.T., Hou T.Y., Hu F., Herr A.B., Strong R.K., Kao C.C., and Li P. 2010. *Structure*. 18, 1032-1043. Copyright (2010) by Elsevier.

Similarly, blunt-ended dsRNA rather than dsRNA with 5'- or 3'-overhangs binds and stimulates the activation of RIG-I (66). Although blunt-ended dsRNA with 5'-ppp potently stimulates the activation of RIG-I, 5'-ppp dsRNA with 5'-overhangs does not stimulate the activation of RIG-I effectively (64). Because different forms of RNA, cell lines, and assays were used in these studies, questions remain about the structural features of viral RNA recognized by RIG-I.

Limited proteolysis of RIG-I bound to dsRNA revealed that its CTD (residues 792-925) is responsible for RNA binding (52). Structures of the CTDs of RIG-I, LGP2, and MDA5 have been determined recently, revealing a novel and conserved RNA-binding domain structure (52,60,72,73). The structure of LGP2 CTD bound to dsRNA showed that the RNA binding surface is located on a positively charged surface (74). NMR titrations of RIG-I CTD showed that a similar binding surface is involved in RNA binding by RIG-I (52). Mutagenesis of several positively charged residues on this surface disrupted RNA binding and signaling by RIG-I (52,72). On the basis of the results from these studies, the ppp-binding site of RIG-I was predicted to be located at a large positively charged surface around residues Lys858 and Lys861 (52,72). However, how RIG-I recognizes the 5'-ppp of viral RNA remains to be established.

To elucidate the structural basis of 5'-ppp dsRNA recognition by RIG-I, we expressed the CTD of human RIG-I, conducted extensive binding studies using RNA with different structures, and determined the crystal structures of RIG-I CTD bound to a 14-bp GC-rich and a 12-bp AU-rich dsRNA. The structures revealed that RIG-I CTD recognizes the termini of dsRNA and interacts with the 5'-ppp through extensive

electrostatic interactions. Mutagenesis of key residues at the binding surface affected RNA binding and signaling by RIG-I.

Materials and methods

Construction of RIG-I CTD expression plasmids by molecular cloning

The cDNA of human RIG-I (DDX58) was purchased from Open Biosystems (Thermo Scientific). The region encoding RIG-I CTD (residues 802-925) was cloned into bacterial expression vector pET22b(+) (Novagen) between *Nde*I and *Xho*I restriction sites, in order to be expressed with an C-terminal 6×His-tag. Residue Cys829 was mutated to serine to prevent dimerization of the protein. RIG-I CTD was also cloned into bacterial expression vector pGEX (GE Healthcare) between *Nde*I and *Sal*I restriction sites, to be expressed with an N-terminal GST-tag.

The target genes were amplified by polymerase chain reaction (PCR) using a Platinum *Pfx* DNA Polymerase kit (Invitrogen) according to standard protocols. PCR products were confirmed by agarose gel electrophoresis and purified using a QIAquick PCR purification kit (QIAGEN). DNA primers (Integrated DNA Technologies, IDT) used in this study were:

For pET22b(+)

Forward primer 5'- TATATATACATATGGATAAGGAAAATAAAAAACTGCTC -3'

Reverse primer 5'- TATATACTCGAGTTTGGACATTTCTGCTGG -3'

For pGEX

Forward primer 5'- TATATATACATATGGATAAGGAAAATAAAAAACTG -3'

Reverse primer 5'- TATATAGTCGACTCATTTGGACATTTCTGCTGG -3'

PCR products and vectors were double-digested using restriction endonucleases and corresponding buffers (New England BioLabs Inc., NEB) as recommended by the manufacturer. Digested products were purified from agarose gels using a QIAquick Gel Extraction kit (QIAGEN). For ligation, a three- to ten-fold excess of the digested insert was incubated with linearized vector and T4 DNA ligase (Invitrogen), in a 10-µl reaction according to the manufacturer's instructions. Ligation products were transformed into *Escherichia coli* (*E. coli*) DH5α cells which were selected on Luria Broth (LB) agar plates containing 100 µg/ml ampicilin. Plasmid DNA was isolated using a QIAprep Spin Miniprep kit (QIAGEN). All plasmid constructs have been confirmed by plasmid DNA sequencing (Institute of plant genomics and biotechnology, Texas A&M University).

Site-directed mutagenesis

For native RIG-I CTD construct, residue Cys829 was mutated to serine to prevent dimerization of the protein. For dsRNA binding surface analysis, 11 residues were mutated for binding studies. Point mutations were introduced by PCR using a QuikChange Site-Directed Mutagenesis kit (Stratagene). DNA primers (IDT) used in this study were:

For C829S (used as native RIG-I CTD)

Forward primer 5'- GAGTGATAGAGGAATGCCATTACACTGTGCTTGG -3'

Reverse primer 5'- CCAAGCACAGTGTAATGGCATTCCTCTATCACTC -3'

For H830A

Forward primer

5'- TAAGAGTGATAGAGGAATCCGCTTACACTGTGCTTGGAGATG -3'

Reverse primer

5'- CATCTCCAAGCACAGTGTAAGCGGATTCCTCTATCACTCTTA -3'

For H847E

Forward primer

5'- AGGAATGCTTTGTGAGTAGACCAGAGCCCAAGCCAAAGC -3'

Reverse primer 5'- GCTTTGGCTTGGGCTCTGGTCTACTCACAAAGCATTCCT -3'

For K849E

Forward primer 5'- TGAGTAGACCACATCCCGAGCCAAAGCAGTTTTCA -3'

Reverse primer 5'- TGAAAACTGCTTTGGCTCGGGATGTGGTCTACTCA -3'

For K851E

Forward primer 5'- AGTAGACCACATCCCAAGCCAGAGCAGTTTTCAAG -3'

Reverse primer 5'- CTTGAAAACTGCTCTGGCTTGGGATGTGGTCTACT -3'

For F853S

Forward primer

5'- CCACATCCCAAGCCAAAGCAGAGTTCAAGTTTTGAAAAAAGAGCA -3'

Reverse primer

5'- TGCTCTTTTTTCAAAACTTGAACTCTGCTTTGGCTTGGGATGTGG -3'

For K858E

Forward primer

5'- CCAAAGCAGTTTTCAAGTTTTGAAGAGAGAGCAAAGATATTCTGTG -3'

Reverse primer

5'- CACAGAATATCTTTGCTCTCTCTTCAAAACTTGAAAACTGCTTTGG -3'

For K861E

Forward primer

5'- CAGTTTTCAAGTTTTGAAAAAAGAGCAGAGATATTCTGTGCCCG -3'

Reverse primer

5'- CGGGCACAGAATATCTCTGCTCTTTTTTCAAAACTTGAAAACTG -3'

For D872A

Forward primer 5'- CAGAACTGCAGCCATGCCTGGGGAATCCATGTG -3'

Reverse primer 5'- CACATGGATTCCCCAGGCATGGCTGCAGTTCTG -3'

For K907E

Forward primer 5'- TTCAGACGCTGTACTCGGAGTGGAAGGACTTTCAT -3'

Reverse primer 5'- ATGAAAGTCCTTCCACTCCGAGTACAGCGTCTGAA -3'

For K909E

Forward primer 5'- CGCTGTACTCGAAGTGGGAGGACTTTCATTTTGAG -3'

Reverse primer 5'- CTCAAAATGAAAGTCCTCCCACTTCGAGTACAGCG -3'

Sequences of the mutants were confirmed by plasmid DNA sequencing (Institute of plant genomics and biotechnology, Texas A&M University).

RIG-I CTD expression and purification

For the overexpression of recombinant native RIG-I CTD and various mutants in *E. coli* system, competent BL21(DE3) cells were transformed with plasmid DNA

carrying the gene of interest. Cells were grown at 37°C in LB medium in the presence of antibiotics (100 μg/ml ampicilin) to an OD_{600} of 0.6-0.8. 0.2 mM of $ZnSO_4$ was added to the medium because RIG-I CTD is a Zinc binding protein. Protein expression was induced by the addition of Isopropyl-β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. After further growth overnight at 15°C, cells were harvested by centrifugation at 4°C, 6000 rpm for 10 min.

For purification, cell pellets were resuspended in pre-chilled Tris-NaCl buffer (150 mM NaCl, and 20 mM Tris-HCl pH 7.5) and disrupted by sonication at 60% amplitude for 5 to 10 min, using a Sonic Dismembrator (Model 500, Fisher Scientific). Cell lysate was then centrifuged down at 4°C, 8000 rpm for 15 min. Supernatant was further centrifuged at 4°C, 16000 rpm for 20 min to completely remove the cell debris.

The 6×His-tagged fusion proteins were first purified by nickel affinity chromatography on a His-Select nickel affinity resin (Sigma-Aldrich) column. After loading of soluble extracts, the resin was washed with Washing Buffer (500 mM NaCl, 20 mM Tris-HCl pH 7.5, and 10 mM imidazole) to remove nonspecific proteins, prior to elution with Elution Buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.5, and 250 mM imidazole). The purified mutants of RIG-I CTD were directly used for binding studies. For crystallization purpose, native RIG-I CTD was further purified by gel filtration chromatography on a Superdex75 (1.6 × 60) column (GE Healthcare) eluted with the Tris-NaCl buffer. The purity of the proteins was analyzed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) on an 18% gel, which was stained with Coomassie Brilliant blue R-250.

The GST-tagged RIG-I CTD was purified by a Glutathione Sepharose 4 Fast Flow (GE Healthcare) column. After loading of soluble extracts, the column was washed with PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4, pH 7.3). The bound protein was eluted by 10 mM reduced glutathione (GSH), and 50 mM Tris-HCl pH 8.0.

Preparation of RNAs used in binding studies and crystallization

Different forms of RNA used in the binding studies and crystallizations are listed in Table 2.1. The non-ppp RNA oligomers were generated by chemical synthesis (IDT). The 5'-ppp RNA were transcribed *in vitro* from DNA template (IDT) using T7 RNA polymerase (expressed and purified from bacterial system in lab).

The transcription reactions were carried out at 37˚C for 3 h, in a reaction mixture containing 40 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 2 mM spermidine, 10 mM dithiothreitol (DTT), 4.8 mM rNTP mix (1.2 mM each), 1 U/ml inorganic pyrophosphatase (IPP), 2 μM dsDNA template, and 0.15 mg/ml T7 RNA polymerase. The transcription products were purified by gel filtration chromatography on a Superdex75 (1.6 × 60) column (GE Healthcare) eluted with the Tris-NaCl buffer, and analyzed by urea-based denaturing PAGE. Each of the purified 5'-ppp RNA showed a single band with the expected length on the denaturing gels.

Table 2.1 Sequences of RNA used in this study.

The 5'-ppp RNA were synthesized by *in vitro* transcription and purified by gel filtration chromatography or denaturing PAGE. The blunt-ended dsRNA were chemically synthesized by IDT.

14-bp GC-rich 5'-ppp dsRNA	5' pppGGCGCGCGCGCC 3' 3' CCGCGCGCGCGGppp 5'
12-bp AU-rich 5'-ppp dsRNA	5' pppAUAUAUAUAUAU 3' 3' UAUAUAUAUAUAppp 5'
14-bp blunt-ended dsRNA without 5'-ppp	5' GGCGCGCGCGCC 3' 3' CCGCGCGCGCGG 5'
13-nt 5'-ppp ssRNA	5' pppGGCGCAUAGGCCG 3'
12-bp 5'-ppp dsRNA with 5'-overhangs	5' pppAUAUGCGCGCGCGC 3' 3' CGCGCGCGCGUAUAppp 5'
12-bp 5'-ppp dsRNA with 3'-overhangs	5' pppGCGCGCGCGCAUAU 3' 3' UAUACGCGCGCGCGppp 5'
24-bp 5'-ppp dsRNA	5' pppGGCGCGCGAGUCGACUCGCGCGCC 3' 3' CCGCGCGCUCAGCUGAGCGCGCGGppp 5'
27-nt 5'-ppp ssRNA	5' pppGGUGCAGAUGAACUUCAGGGUCAGCUU 3'
27-bp blunt-ended dsRNA without 5'-ppp	5' AAGCUGACCCUGAAGUUCAUCUGCACC 3' 3' UUCGACUGGGACUUCAAGUAGACGUGG 5'

All the dsRNA used were generated by annealing a ssRNA with palindromic sequence or two ssRNA strands with complementary sequences. The samples were annealed by heating the ssRNA at 95°C for 5 min and cooling gradually to room temperature.

RNA binding studies by gel filtration chromatography

In the binding studies of RIG-I CTD with different forms of RNA by gel filtration chromatography, each of the RNA samples (at about 100 μM) was mixed with equal volume of RIG-I CTD (at about 200μM), and 100 μl of samples were injected over a Superdex200 (10/300 GL) column (GE Healthcare) eluted with the Tris-NaCl buffer. The column was calibrated with a set of protein standards of different sizes (Bio-Rad) to allow the estimation of the sizes of the protein and RNA complexes.

Binding studies for the mutants with the three different forms of RNA were carried out in the same way, but with final concentrations of about 36:12 μM (protein:RNA).

RNA binding studies by electrophoretic mobility shift assay (EMSA)

In the binding studies by EMSA, each type of RNA was mixed with excess RIG-I CTD or mutants at a molar ratio of 1:3 at final concentrations of about 1.6:5 μM. The mixtures were resolved on 12% native polyacrylamide gels in Tris-boric acid buffer (89 mM Tris, 89 mM boric acid) without EDTA. The gels were stained with ethidium bromide (EtBr) and visualized using a Gel Dox XR (Bio-Rad) system.

Analytical ultracentrifugation (AUC)

To determine the stoichiometry of RIG-I CTD binding to the 14-bp 5'-ppp dsRNA, purified RIG-I CTD complex with 14-bp dsRNA was analyzed by AUC. The experiments were carried out by our collaborators at the University of Cincinnati (Dr. Andrew B. Herr group, Department of Molecular Genetics, Biochemistry, and Microbiology).

For sedimentation velocity experiments, 400 µl of samples in Tris-NaCl buffer with 10 mM β-mercaptoethanol were centrifuged overnight at 48000 rpm at 10 °C in a Beckman XL-I using absorbance optics at 304 nm. The data were analyzed by the program Sedfit using the c(s) and c(M) models to determine differential sedimentation coefficient and apparent mass distributions, respectively. The complex sedimented as a single peak, with a sedimentation coefficient of 2.72 S and an estimated molecular mass of 38 kDa, similar to that expected for a 2:1 (protein:RNA) complex (38.5 kDa).

To confirm the stoichiometry, the complex was also analyzed in a sedimentation equilibrium experiment. Briefly, 100 µl of samples of the complex at concentrations of 15, 50, and 150 µM were spun at 19000, 22500, and 33000 rpm at 10°C until equilibrium was reached, and were scanned at 295 and 306 nm. The data were trimmed using WinReedit and globally analyzed using WinNonlin. The global analysis of nine data sets revealed that the data could be described by a single species with a reduced buoyant molecular weight of 1.883 (at a speed of 19000 rpm). To convert to experimental molecular weight, the partial specific volume must be known; for a protein and RNA complex this value corresponds to the weight average of each component in

the sedimenting species. The partial specific volume for RIG-I CTD alone was calculated to be 0.7258 ml/g at 10°C using Sednterp. The partial specific volume for the 14-bp dsRNA was calculated to be 0.5688 ml/g using the NucProt Calculator server (http://www.molmovdb.org/cgi-bin/psv.cgi) and corrected to 10°C in Sednterp. Both 2:1 and 2:2 complexes were considered, with weight-average partial specific volumes of 0.6956 or 0.6716 ml/g, respectively. The molecular weight values calculated from the reduced buoyant molecular weight and appropriate partial specific volumes were only self-consistent in the case of a 2:1 complex, yielding an experimental molecular weight of 37339 Da (95% confidence interval, 33882 – 40796 Da).

RNA binding studies by surface plasmon resonance (SPR)

SPR experiments were carried out by our collaborators at the Fred Hutchinson Cancer Research Center in Seattle (Dr. Roland K. Strong group, Division of Basic Sciences).

Binding interactions between RIG-I CTD and three different forms of RNA were analyzed by SPR at 25°C in HBS-EP$^+$ buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, and 0.05% v/v P-20 surfactant) on a Biacore T100 system (Biacore AB, GE Healthcare). The goat anti-GST monoclonal antibody (mAb) (Biacore AB) at 30 μg/ml in 10 mM sodium acetate (pH 5.0) was immobilized on a CM5 sensor chip (Biacore AB) by standard amine coupling chemistry. Immobilization of 13000 response units (RU) of the antibody resulted in optimal responses in subsequent analyses. GST fusion protein of

RIG-I CTD (at 5 μg/ml) was captured over the immobilized anti-GST mAb surface at a flow rate of 10 μl/min for 120 seconds to reach 1200 RU response.

For equilibrium binding studies, dilution series of different forms of RNA were injected in randomized duplicate runs at flow rates of 30, 50, and 50 μl/min for 10, 7, and 5 min for the 14-bp 5'-ppp dsRNA, the 14-bp blunt-ended dsRNA, and the 13-nt 5'-ppp ssRNA, respectively, followed by a 5-min dissociation phase. Concentrations for the three forms of RNA in the dilution series were 0.08-10 nM, 0.6-80 nM, and 2.5-160 nM, respectively.

To study the kinetics of RNA binding by RIG-I CTD, dilution series of the three different forms of RNA were injected in randomized duplicate runs at a flow rate of 50 μl/min for 5 or 7 min, followed by a 3 or 4 min dissociation phase. Concentrations for the three forms of RNA in the dilution series were 0.3-5 nM, 0.6-10 nM, and 2.5-40 nM, respectively. Optimal regeneration of the sensor chip was achieved by injection of 10 mM glycine (pH 2.2) at a flow rate of 20 μl/min for 210 seconds followed by HBS-EP$^+$ buffer stabilization for 5 min. Sensorgrams obtained from the SPR measurements were analyzed by the double-subtraction method described by Myszka (91). The signal from the reference flow cell was subtracted from the analyte binding response obtained from the flow cell with captured ligand. Buffer reference responses were then averaged from multiple injections. The averaged buffer reference response was then subtracted from analyte binding responses, and the final double-referenced data were analyzed with BIAevaluation 3.0 software (GE Healthcare). For the equilibrium binding data, steady-state binding levels of the analytes were plotted against analyte concentration, from

which the equilibrium binding constant was estimated. For the kinetic binding data, a one to one binding model was used to globally fit the data to derive the association and dissociation rate constants.

IFN-β luciferase reporter gene assays

IFN-β luciferase reporter gene assays were used to analyze the activities of three different forms of RNA in stimulating RIG-I signaling in the human embryonic kidney (HEK) 293T cell line. These cell-based assays were performed by our collaborators at the Indiana University, Bloomington (Dr. C. Cheng Kao group, Department of Molecular and Cellular Biochemistry). The RNA used in the assays were purified either by preparative denaturing PAGE or by gel filtration chromatography.

The luciferase assays used actively growing HEK 293T cells plated in CoStar White 96-well plates at 4.4×10^4 cells per well. When the cells were ~80% confluent, they were transfected with a mixture of Lipofectamine 2000 reagent (Invitrogen) and constant amounts of the luciferase reporter plasmid IFN-β luc (30 ng per transfection; gift from Dr. Rongtuan Lin, Lady Davis Institute for Medical Research, Montreal, QC, Canada), the *Renilla* luciferase control plasmid phRL-TK (5 ng; Promega), and human RIG-I plasmid pUNO-hRIG-I (0.5 ng; Invivogen). The cells were incubated for 24 h to allow expression from the plasmids. Three different forms of RNA at 5, 10 and 50 nM concentrations were then transfected into the cells to stimulate RIG-I dependent signaling. After 16 h incubation, the cells were analyzed using the Dual-Glo Luciferase Assay System reagents (Promega), quantifying luminescence with the BioTek Synergy2

Plate Reader. The ratios of firefly luciferase reporter over *Renilla* luciferase control were determined.

Signaling of RIG-I mutants were analyzed under similar conditions. Mutations of full-length RIG-I were performed using pUNO-hRIG-I (Invivogen) as template. The cells were transfected with plasmid of RIG-I wild type or mutants along with IFN-β luciferase reporter and Renilla luciferase control plasmids, and stimulated with three forms of RNA at a final concentration of 50 nM. The firefly over *Renilla* luciferase ratio obtained with wild-type RIG-I induced with RNA was considered as 100% activity. Expression of wild-type and mutants of full-length RIG-I in the transfected cells were confirmed by western blot using anti-RIG-I antibody (C-15, sc-48929, Santa Cruz biotechnology).

Purification and Crystallization of RIG-I CTD and dsRNA complexes

Purified RIG-I CTD was mixed with the 14-bp or the 12-bp 5'-ppp dsRNA at a molar ratio of about 3:1 (excess protein), and the 2:1 (RIG-I CTD:dsRNA) complexes were purified by gel filtration chromatography on a Superdex75 (1.6 x 60) column (GE Healthcare) eluted with the Tris-NaCl buffer. After purification, 5 mM DTT was added to the complexes, and the samples were concentrated for crystallization.

The complex of RIG-I CTD with the 14-bp GC-rich dsRNA was crystallized at about 15 mg/ml, in 18% polyethylene glycol (PEG) 3350, 0.2 M $(NH_4)_2SO_4$, and 0.1 M Bis-Tris (pH 5.5) at 4°C. Large single crystals grew in about one week.

The complex of RIG-I CTD with the 12-bp AU-rich dsRNA was crystallized under similar conditions in 18% PEG 3350, 0.2 M $(NH_4)_2SO_4$, and 0.1 M Bis-Tris (pH 6.5).

Data collection and structural determination

For data collection, the crystals were transferred stepwise into a cryoprotectant containing 30% PEG 400 and were flash frozen in liquid nitrogen. Diffraction data were collected using a Rigaku RAXIS IV^{++} image plate detector mounted on a Rigaku Micromax-007HF generator. The diffraction data were processed with the HKL2000 package (92).

The complex of the 14-bp GC-rich dsRNA and RIG-I CTD crystallizes in orthorhombic space group $P2_12_12_1$ with the following cell dimensions: a = 41.49 Å, b = 96.87 Å, c = 103.40 Å. The crystallographic asymmetric unit (ASU) contains one 2:1 (RIG-I CTD:dsRNA) complex. The 12-bp AU-rich dsRNA complex crystallizes in hexagonal space group $P6_5$ with cell dimensions as follows: $a = b$ = 82.90 Å, c = 111.16 Å, and $\gamma = 120°$. There is one 2:1 (RIG-I CTD:dsRNA) complex in the ASU. The crystal is hemihedrally twinned with a twin fraction of 0.478 with twinning operators h, $-h-k$, $-l$. The twining fraction was determined using CNS (93).

Structures of the RIG-I CTD in complex with the 14-bp and 12-bp dsRNA were determined by molecular replacement using MOLREP in the CCP4 suite (94) using the crystal structure of RIG-I CTD in isolation as the search model (PDB code: 2QFB, chain A). The structural models were rebuilt using O. After several rounds of rebuilding and

refinement with CNS, electron density for the dsRNA became apparent. A 14-bp and a 12-bp dsRNA were manually docked into the electron density map and rebuilt with O (95). The structures were refined by several rounds of positional, simulated annealing and individual B-factor refinement using CNS followed by extensive manual remodeling. The structure of RIG-I CTD bound to the 12-bp AU-rich dsRNA was refined in CNS against the twinned data. A total of 20357 reflections, which correspond to 86% of a complete dataset, were used in the structural refinement. Only 5.5% of the diffraction data (1309 reflections) were rejected due to twinning. Noncrystallographic symmetry (NCS) restraints were not applied during the refinement because of the different crystal packing environment of the RIG-I CTDs in the two complexes.

The atomic coordinates and structure factors of RIG-I CTD bound to the 14-bp and the 12-bp 5'-ppp dsRNA have been deposited with the RCSB Protein Data Bank under the accession codes 3LRN and 3LRR. The structural figures were generated with PyMol (http:///www.pymol.org).

Results

RIG-I CTD interacts with various forms of RNA

Understanding of RIG-I structure and function requires the elucidation of the chemical basis for viral RNA recognition by RIG-I. Since the CTD of RIG-I is responsible for viral RNA recognition, we have expressed and purified human RIG-I CTD and characterized its ligand binding properties with various forms of RNA, by gel filtration chromatography.

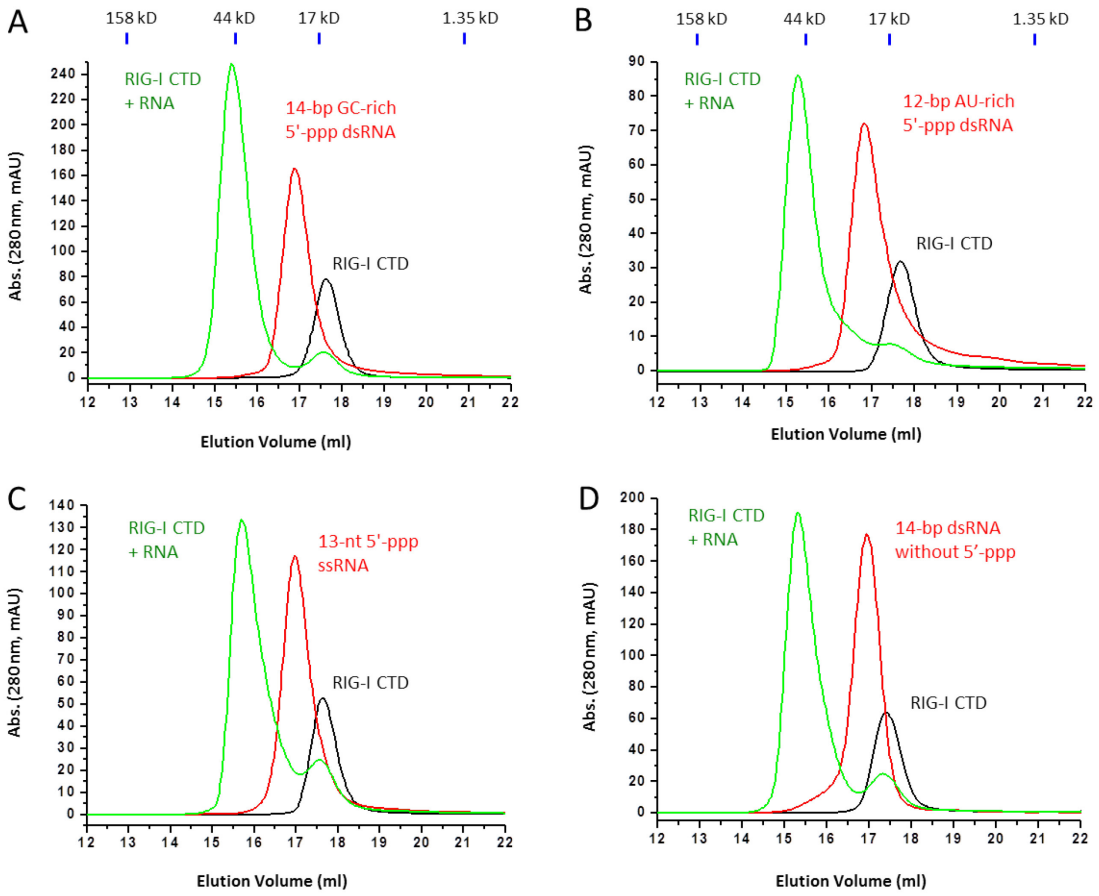


Figure 2.1 Bind studies of RIG-I CTD and RNA with or without 5'-ppp groups by gel filtration chromatography.

(A) RIG-I CTD binds to a 14-bp GC-rich 5'-ppp dsRNA. RIG-I CTD is shown by the black chromatogram. The RNA is shown by the red chromatogram. The mixture of the protein and RNA is shown by green. Elution volumes of four protein standards are shown above the chromatograms.

(B) RIG-I CTD binds to a 12-bp AU-rich 5'-ppp dsRNA.

(C) RIG-I CTD binds to a 13-nt 5'-ppp ssRNA.

(D) RIG-I CTD binds to a 14-bp GC-rich dsRNA without a 5'-ppp group.

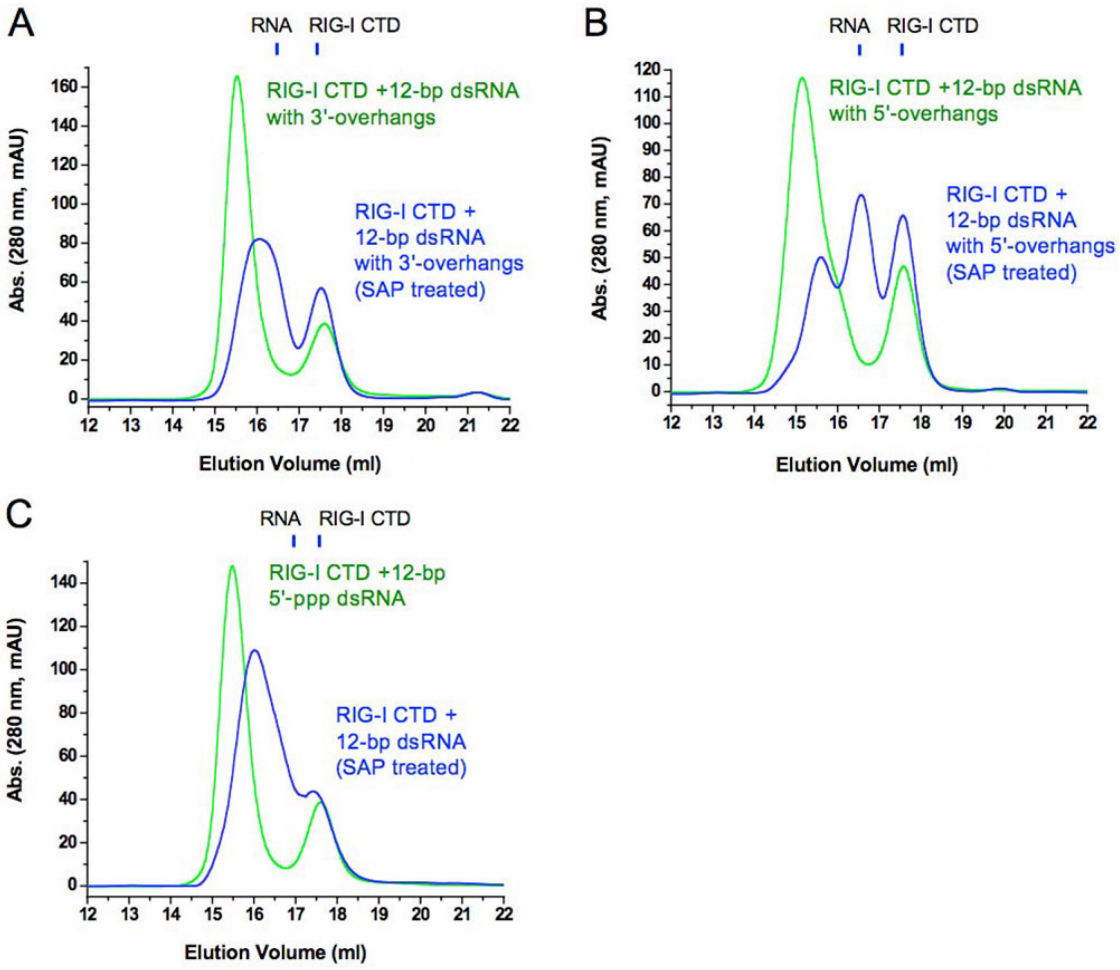


Figure 2.2 Bind studies of RIG-I CTD and RNA by gel filtration chromatography. (A) RIG-I CTD binds to a 12-bp 5'-ppp dsRNA with 3'-overhangs (green). Treatment of the dsRNA with SAP reduced the binding between RIG-I CTD and the RNA (blue). Positions of the protein and the RNA alone are shown above the chromatograms. (B) RIG-I CTD binds to a 12-bp 5'-ppp dsRNA with 5'-overhangs (green). Treatment of the dsRNA with SAP reduced the binding between RIG-I CTD and the RNA (blue). (C) Treatment of a blunt-ended 12-bp 5'-ppp dsRNA (green) with SAP reduced its binding for RIG-I CTD (blue).

First, we tested the bindings of blunt-ended RNA with 5'-ppp groups. A 14-bp GC-rich dsRNA, a 12-bp AU-rich dsRNA and a 13-nt ssRNA were synthesized by *in vitro* transcription using T7 RNA polymerase (Table 2.1). We found that the 14-bp dsRNA with 5'-ppp binds to RIG-I CTD and forms a stable complex which is represented as a single peak eluted at higher molecular weight (Figure 2.1A). The 12-bp AU-rich dsRNA also binds to RIG-I CTD and forms a complex (Figure 2.1B), indicating that RIG-I has no preference for specific bases within the RNA. Surprisingly, the 13-nt ssRNA with 5'-ppp also binds to RIG-I CTD (Figure 2.1C). In addition, we found a blunt-ended dsRNA with the same sequence as the 14-bp GC-rich 5'-ppp dsRNA but lacking the triphosphate group binds to RIG-I CTD as well (Figure 2.1D).

Second, we tested whether the 5'- or 3'-end overhanging regions of 5'-ppp dsRNA affect RNA binding to RIG-I CTD. We found that 5'-ppp dsRNA with either 3'- or 5'-overhanging nucleotides retained binding to RIG-I CTD (Table 2.1, Figure 2.2A and B). However, treatment of these dsRNA with shrimp alkaline phosphatase (SAP) that removes the 5'-ppp groups dramatically reduced their binding with RIG-I CTD (Figure 2.2A and B), demonstrating that the 5'-ppp increases the affinity of the dsRNA for RIG-I. Consistent with this observation, treatment of the 12-bp 5'-ppp dsRNA without 3'- or 5'-overhangs using SAP also reduced its binding for RIG-I CTD (Figure 2.2C).

The stoicheometry between RIG-I CTD and RNA ligands

From the gel filtration chromatograms of RIG-I CTD binding study, the molecular weight of the protein and RNA complexes, and hence the stoicheometry between them, can be calculated by comparing the elution volumes of the complexes with those of the protein standards. The binding ratio of RIG-I CTD with dsRNA, with or without the 5'-ppp group, is estimated to be about 2:1, while that of RIG-I CTD with 5'-ppp ssRNA is 1:1.

To confirm the stoichiometry estimated from the gel filtration binding studies, sedimentation velocity and equilibrium analyses by AUC were performed with the help of our collaborators. The analyses of purified RIG-I CTD bound to the 14-bp GC-rich 5'-ppp dsRNA yield an experimental molecular weight of about 38 kDa, which agrees with a 2:1 complex (RIG-I CTD:dsRNA; 38.5 kDa), demonstrating that the stoichiometry between the protein and the dsRNA is 2:1 (Figure 2.3).

The RNA binding affinities of RIG-I CTD

The affinity and kinetic properties of ligand binding to immunoreceptors determine the strength and duration of the signals initiated by these receptors. To determine the affinities and kinetic properties of RNA binding by RIG-I CTD, we expressed and purified GST-tagged RIG-I CTD and conducted RNA binding studies by SPR with the help of our collaborators.

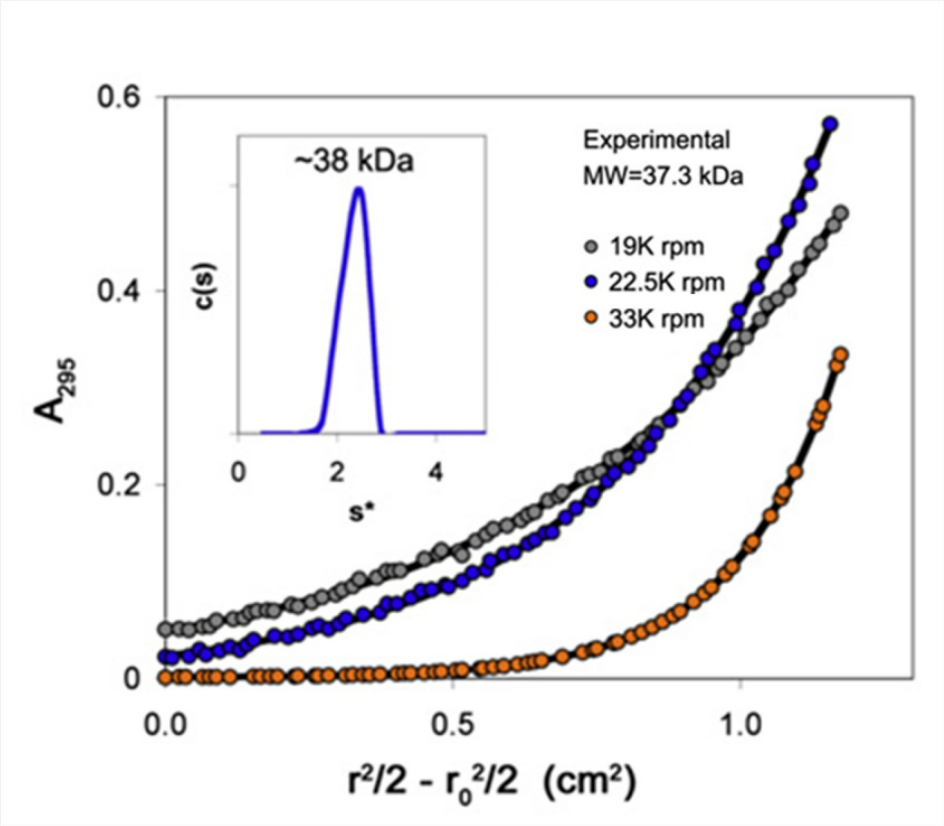
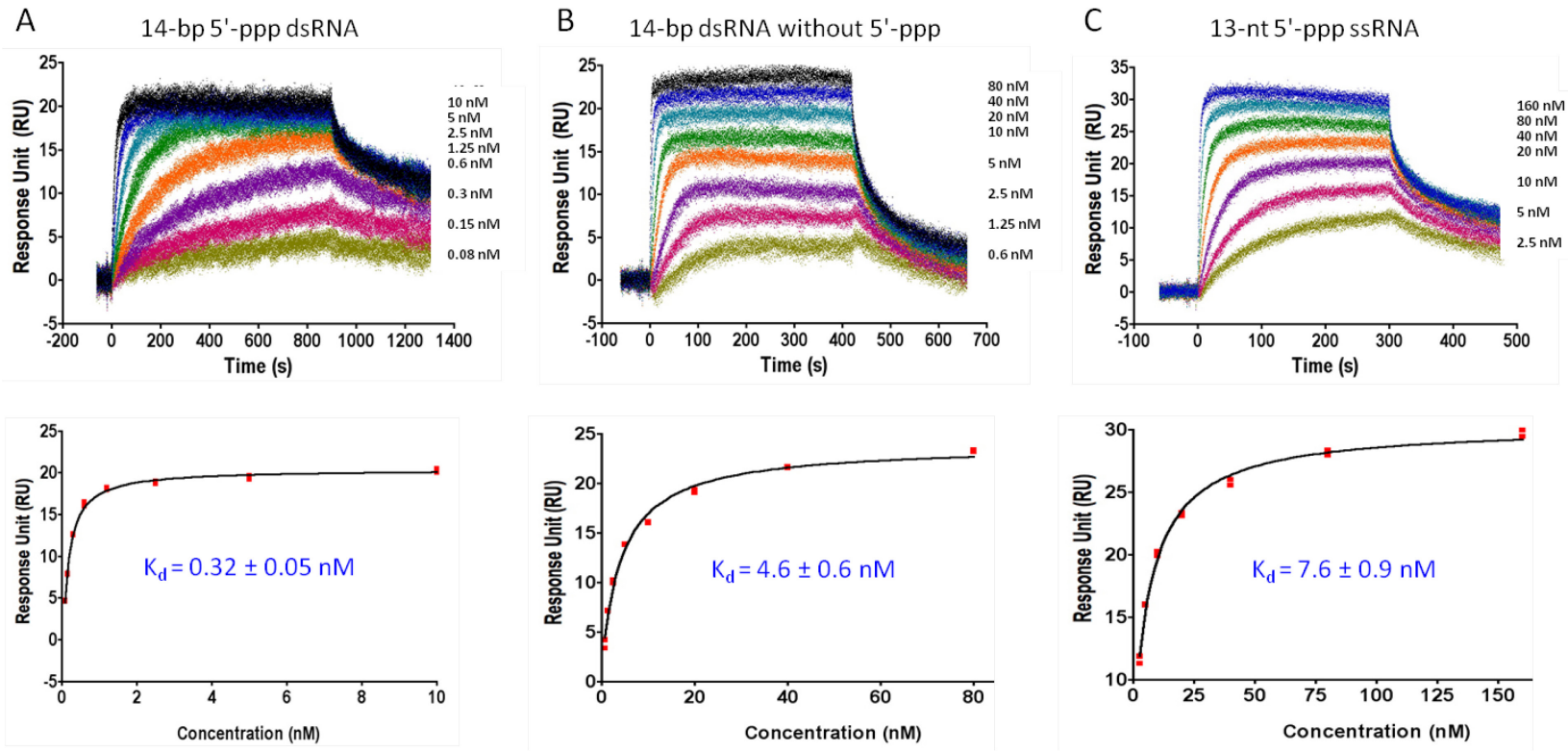
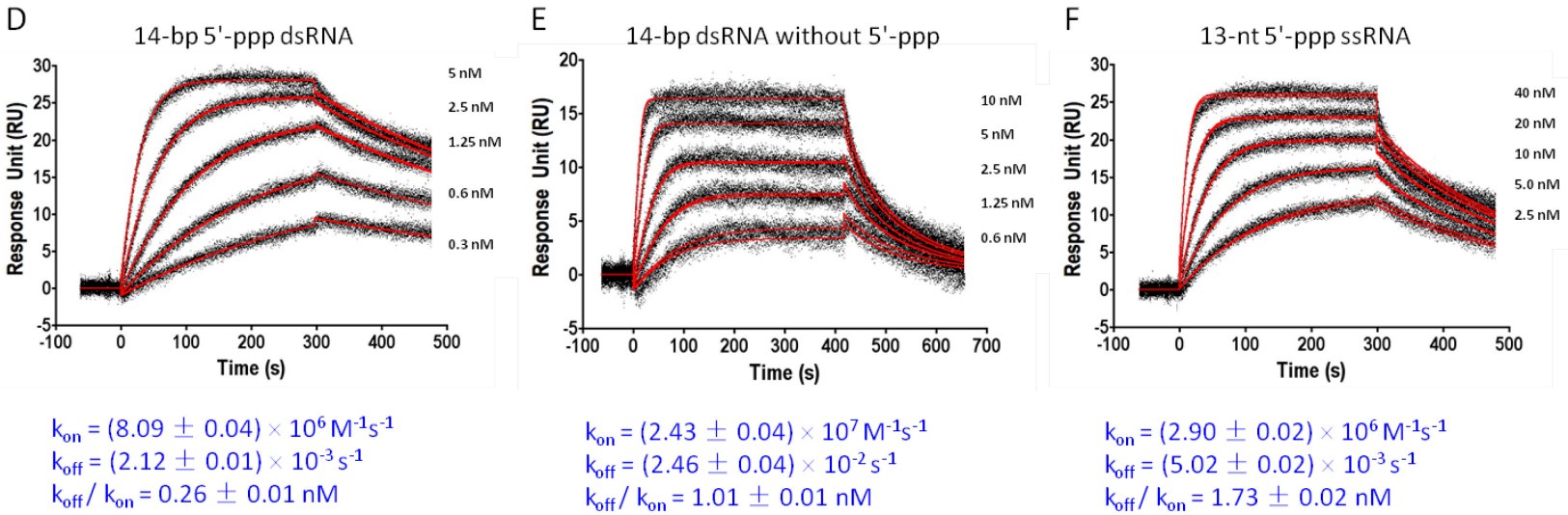


Figure 2.3 Stoicheometry between RIG-I CTD and the 14-bp GC-rich 5'-ppp dsRNA determined by AUC.

AUC analyses of a purified protein complex by sedimentation velocity (inset) and equilibrium. The experimental molecular weight values from velocity (about 38 kDa) and equilibrium (37339 Da) data were consistent only with the formation of a 2:1 complex.

Figure 2.4 The binding affinities between RIG-I CTD and RNA determined by SPR.
(A) Equilibrium binding study of RIG-I CTD with the 14-bp GC-rich 5'-ppp dsRNA. The dissociation constant (K_d) was derived by fitting of the equilibrium binding data to a one site binding model (lower panels).
(B) Equilibrium binding study of RIG-I CTD with the 14-bp non-ppp dsRNA.
(C) Equilibrium binding study of RIG-I CTD with a 13-nt 5'-ppp ssRNA.
(D) Kinetic binding studies of RIG-I CTD with the 14-bp GC-rich 5'-ppp dsRNA. The association and dissociation rate constants (k_{on} and k_{off}) were derived from global fitting of the binding data to a 1:1 binding model (red curves). Half-life ($t_{1/2}$) for each of the protein:RNA complex was calculated from the k_{off}. The calculated dissociation constant ($K_{D,\ calc}$) was derived by division of the dissociation rate constant (k_{off}) by the association rate constant (k_{on}).
(E) Kinetic binding studies of RIG-I CTD with the 14-bp non-ppp dsRNA.
(F) Kinetic binding studies of RIG-I CTD with the 13-nt 5'-ppp ssRNA.


A
14-bp 5'-ppp dsRNA
Response Unit (RU)
Time (s)
10 nM
5 nM
2.5 nM
1.25 nM
0.6 nM
0.3 nM
0.15 nM
0.08 nM
B
14-bp dsRNA without 5'-ppp
Response Unit (RU)
Time (s)
80 nM
40 nM
20 nM
10 nM
5 nM
2.5 nM
1.25 nM
0.6 nM
C
13-nt 5'-ppp ssRNA
Response Unit (RU)
Time (s)
160 nM
80 nM
40 nM
20 nM
10 nM
5 nM
2.5 nM
Response Unit (RU)
Kd = 0.32 ± 0.05 nM
Concentration (nM)
Response Unit (RU)
Kd = 4.6 ± 0.6 nM
Concentration (nM)
Response Unit (RU)
Kd = 7.6 ± 0.9 nM
Concentration (nM)



$k_{on} = (8.09 \pm 0.04) \times 10^{6}\ M^{-1}s^{-1}$
$k_{off} = (2.12 \pm 0.01) \times 10^{-3}\ s^{-1}$
$k_{off} / k_{on} = 0.26 \pm 0.01$ nM

$k_{on} = (2.43 \pm 0.04) \times 10^{7}\ M^{-1}s^{-1}$
$k_{off} = (2.46 \pm 0.04) \times 10^{-2}\ s^{-1}$
$k_{off} / k_{on} = 1.01 \pm 0.01$ nM

$k_{on} = (2.90 \pm 0.02) \times 10^{6}\ M^{-1}s^{-1}$
$k_{off} = (5.02 \pm 0.02) \times 10^{-3}\ s^{-1}$
$k_{off} / k_{on} = 1.73 \pm 0.02$ nM

Figure 2.4 Continued.

Equilibrium binding studies showed that the affinity of RIG-I CTD for the 14-bp 5'-ppp dsRNA is 0.32 ± 0.05 nM (Figure 2.4A). The affinity of RIG-I CTD for the 14-bp dsRNA without 5'-ppp is 4.6 ± 0.6 nM (Figure 2.4B), about 15 times lower than the affinity for the 5'-ppp dsRNA. The 13-nt 5'-ppp ssRNA binds to RIG-I CTD with an affinity of 7.6 ± 0.9 nM (Figure 2.4C). The calculated dissociation constants ($K_{D,\ calc} = k_{off} / k_{on}$) derived from the kinetic binding studies are similar to those derived from the equilibrium binding studies (Figure 2.4). Although RIG-I CTD interacts with RNA primarily through electrostatic interactions, binding of the 14-bp 5'-ppp dsRNA by RIG-I CTD shows slow association rate and slow dissociation rate (Figure 2.4D), suggesting that a conformational change of the protein is likely needed for RNA binding. It is possible that other types of interactions, such as hydrogen bonding and hydrophobic interactions may also contribute to RNA binding. In contrast, the binding of blunt-ended dsRNA lacking a 5'-ppp shows faster association and dissociation rates (Figure 2.4E). Binding of the 13-nt 5'-ppp ssRNA shows kinetic properties similar to that of the 5'-ppp dsRNA (Figure 2.4F). These results indicate that the 5'-ppp plays an important role in determining the kinetic properties of RNA binding by RIG-I CTD. The higher affinity of the 5'-ppp dsRNA and the longer half-life ($t_{1/2}$ = 327 ± 2 s) of its complex with RIG-I CTD suggest that dsRNA with 5'-ppp should be a more potent stimulator of RIG-I.

RIG-I signaling activities stimulated by various forms of RNA in vivo

The activities of three different forms of RNA in stimulating RIG-I signaling were examined by IFN-β luciferase reporter gene assays. HEK 293T cells were

transfected with plasmid of RIG-I along with IFN-β luciferase reporter and Renilla luciferase control plasmids, before stimulated with 5'-ppp ssRNA, 5'-ppp dsRNA, or dsRNA without 5'-ppp. These studies showed that the three forms of RNA are all capable of activating RIG-I-dependent reporter activities (Figure 2.5). The relative activities of 5'-ppp RNA are similar, whereas the non-ppp dsRNA is slightly less active. When used at the lowest concentration (5 nM), 5'-ppp dsRNA shows higher activity than non-ppp dsRNA or 5'-ppp ssRNA.

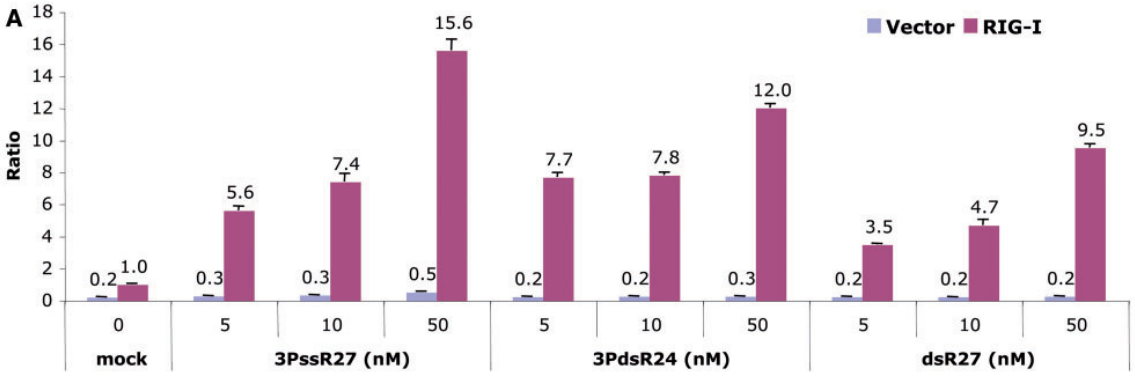


Figure 2.5 The signaling activities of three different RNA ligands for RIG-I by IFN-β luciferase assays.

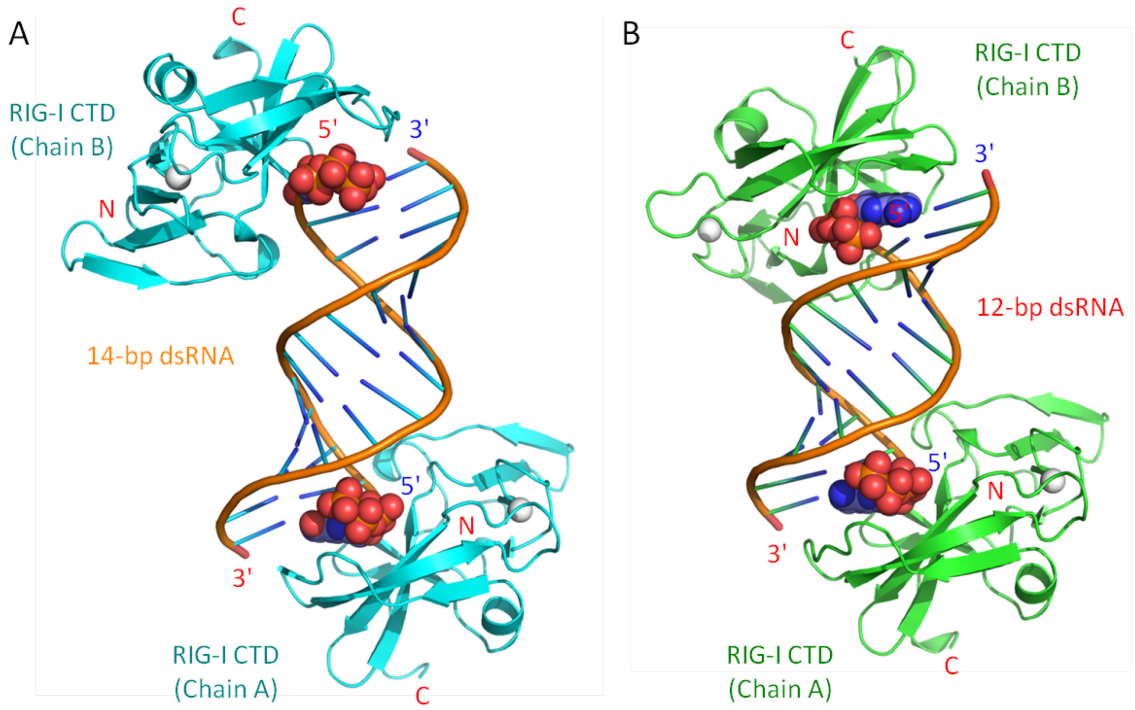
The assays show the activities of a 27-nt 5'-ppp ssRNA (3PssR27), a 24-bp 5'-ppp dsRNA (3PdsR24), and a 27-bp dsRNA without 5'-ppp (dsR27) in stimulating RIG-I signaling in HEK 293T cells. The error bars correspond to the standard deviations of signals from three independent transfections.

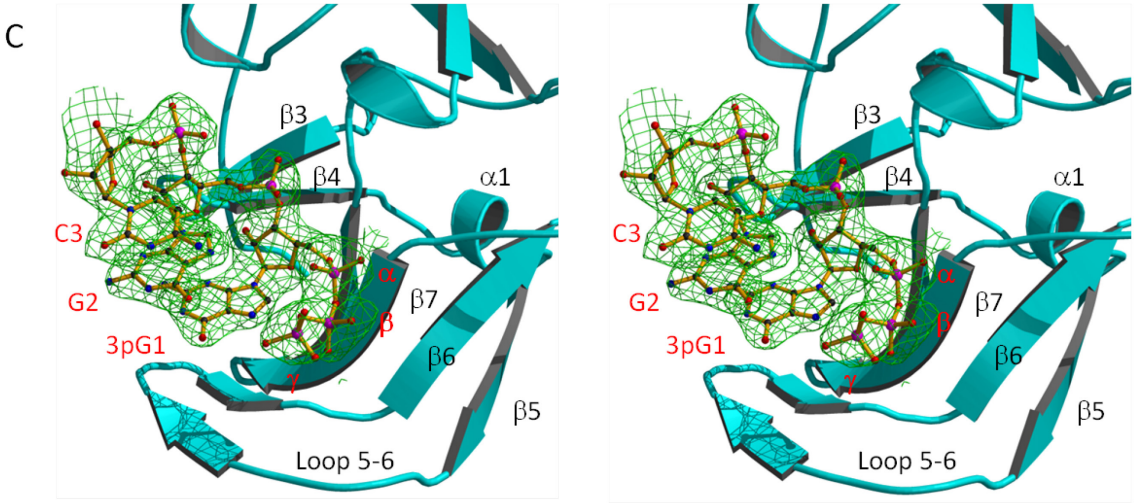
Table 2.2 Data collection and refinement statistics of 5'-ppp dsRNA complex crystals.

	RIG-I CTD : 14-bp dsRNA	**RIG-I CTD : 12-bp dsRNA**
Data Collection		
Space group	$P2_12_12_1$	$P6_5$
Cell dimensions *a*, *b*, *c* (Å) α, β, γ (°)	 41.49, 96.87, 103.40 90.0, 90.0, 90.0	 82.90, 82.90, 111.16 90.0, 90.0, 120.0
Resolution (Å)	50-2.60 (2.69-2.60)[a]	50.00-2.15 (2.23-2.15)[a]
R_{sym} (%)	8.6 (58.7)	7.1 (45.7)
$I/\sigma I$	27.4 (3.2)	36.0 (4.4)
Completeness (%)	98.7 (96.3)	93.5 (90.4)
Redundancy	4.5 (4.4)	6.5 (6.3)
Refinement		
Resolution (Å)	50-2.60	50-2.15
No. of reflections	12530	20357
R_{work}/R_{free} (%)	22.1/28.5	18.7/22.7
No. of atoms Protein RNA Ion Water	 1998 620 2 54	 1998 522 2 0
B-factor Protein RNA Ligand/ion Water	 45.9 69.5 45.0 49.0	 37.3 70.7 35.1 N/A
R.m.s. deviations Bond lengths (Å) Bond angles (°)	 0.013 1.69	 0.009 1.54

[a] Diffraction data were collected using one single crystal for each structure. Values in the parentheses are for the highest-resolution shell.

Figure 2.6 Structures of RIG-I CTD bound to 5'-ppp dsRNA.
(A) Crystal structure of RIG-I CTD bound to the 14-bp GC-rich 5'-ppp dsRNA (PDB code 3LRN). The nucleotides at the 5'-ends of the dsRNA containing the ppp are shown as space filling models. The zinc ions bound to RIG-I CTD are shown as gray spheres.
(B) Structure of the RIG-I CTD bound to the 12-bp AU-rich 5'-ppp dsRNA (PDB code 3LRR). The orientations of RIG-I CTD (Chain A) relative to the dsRNA are the same in the two structures.
(C) Stereo representation of the σ_A weighted $2F_o$ - F_c map for the 5' three nucleotides in the 14-bp GC-rich 5'-ppp dsRNA complex structure. The first nucleotide 3pG1 contains a ppp group. RIG-I CTD is shown as cyan ribbons. The nucleotides are shown in ball-and stick models.


A
C
RIG-I CTD
(Chain B)
5'
3'
N
14-bp dsRNA
5'
N
3'
RIG-I CTD
(Chain A)
C
B
C
RIG-I CTD
(Chain B)
3'
N
12-bp dsRNA
5'
N
3'
RIG-I CTD
(Chain A)
C


C
β3
β4
α1
C3
G2
3pG1
α
β
γ
β7
β6
β5
Loop 5-6
β3
β4
α1
C3
G2
3pG1
α
β
γ
β7
β6
β5
Loop 5-6

Crystal structures of RIG-I CTD bound to 5'-ppp dsRNA

To elucidate the structural basis of 5'-ppp dsRNA recognition by RIG-I, we cocrystallized RIG-I CTD (residues 803 to 923) with a 14-bp GC-rich dsRNA (with self-complementary sequence of 5' pppGGCGCGCGCGCGCC 3') or a 12-bp AU-rich dsRNA (with self-complementary sequence of 5' pppAUAUAUAUAUAU 3'). The crystal structures of the complexes were determined by molecular replacement and refined at 2.60 Å and 2.15 Å resolutions, respectively. Crystals of the 12-bp AU-rich dsRNA complex are merohedrally twinned and the structure was refined against the twinned data using CNS (93). Statistics of data collection and structural refinement are listed in Table 2.2.

The overall structures of the two crystals are shown in Figure 2.6. Consistent with the stoichiometry determined by AUC, the crystallographic ASUs for the two complex crystals contain a 2:1 complex between RIG-I CTD and the dsRNA (Figure 2.6A and B). The complexes exhibit pseudo two-fold symmetry along the two-fold axis of the palindromic dsRNA. The orientations of the dsRNA relative to the RIG-I CTD in the two complexes are similar (Figure 2.6A and B). The 5'-ppp groups of the dsRNA are well defined at the four binding sites in the two complexes and interact with the protein in similar ways (Figure 2.6C). Structures of four RIG-I CTDs in the two complexes are similar; the root mean square deviations (r.m.s.d) between the Cα atoms of RIG-I CTD within each complex or between the two complexes are only 0.3 to 0.5 Å. Discussions of the complex structure below are based on the structure of RIG-I CTD (Chain A) in the 14-bp GC-rich dsRNA complex because of the better-defined electron density map at

this interface (Figure 2.6C) and more complete diffraction data used in the structural refinement (Table 2.2).

The overall structure of RIG-I CTD in the complex is similar to the structure of RIG-I CTD in isolation (r.m.s.d of about 0.7 Å). The largest structural difference between free and RNA bound RIG-I CTD occurs at the loop between strands β5 and β6 (loop 5-6) from His847 to Ser855 (Figure 2.7). The average deviation for the Cα atoms of these eight residues is 1.67 Å in the two structures. As reflected in the nine structures of RIG-I CTD in the crystallographic ASU, this loop is highly flexible in the structures of RIG-I CTD in isolation (72). The structure of RIG-I CTD in solution also revealed that this loop is flexible (52). However, the four structures of the RIG-I CTD in the two complexes are highly similar, suggesting that dsRNA binding stabilizes the specific conformation of loop 5-6.

RIG-I CTD recognizes the termini of 5'-ppp dsRNA

As predicted from our previous structural studies of the LGP2 CTD: dsRNA complex (74), RIG-I CTD recognizes the termini of the 5'-ppp dsRNA (Figure 2.6). The dsRNA in the complex adopts a typical A-form double helical structure. The RNA-binding surface of RIG-I CTD is located at a large saddle-shaped surface defined by the hairpin containing strands β3 and β4, the loop connecting strands β5 and β6, the β-sheet containing strands β5 to β8, the loop connecting strands β8 and β9, and the loop connecting strand β10 to the C-terminus (Figure 2.8).

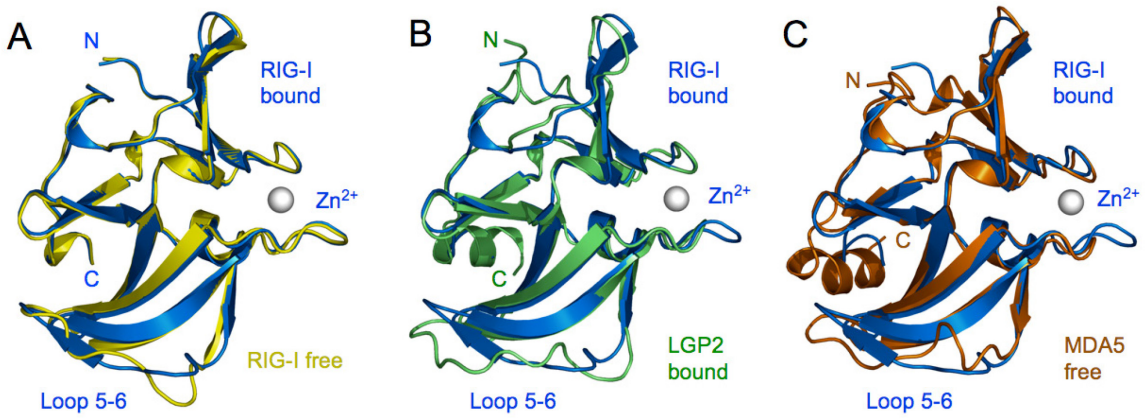


Figure 2.7 The conserved structures of RLR CTDs.

(A) Superposition of RIG-I CTD structures in isolation (yellow) and in complex with 5'-ppp dsRNA (blue).

(B) Superposition of RNA bound RIG-I CTD (blue) and LGP2 CTD (green) structures.

(C) Superposition of structures of RNA bound RIG-I CTD (blue) and MDA5 CTD in isolation (orange).

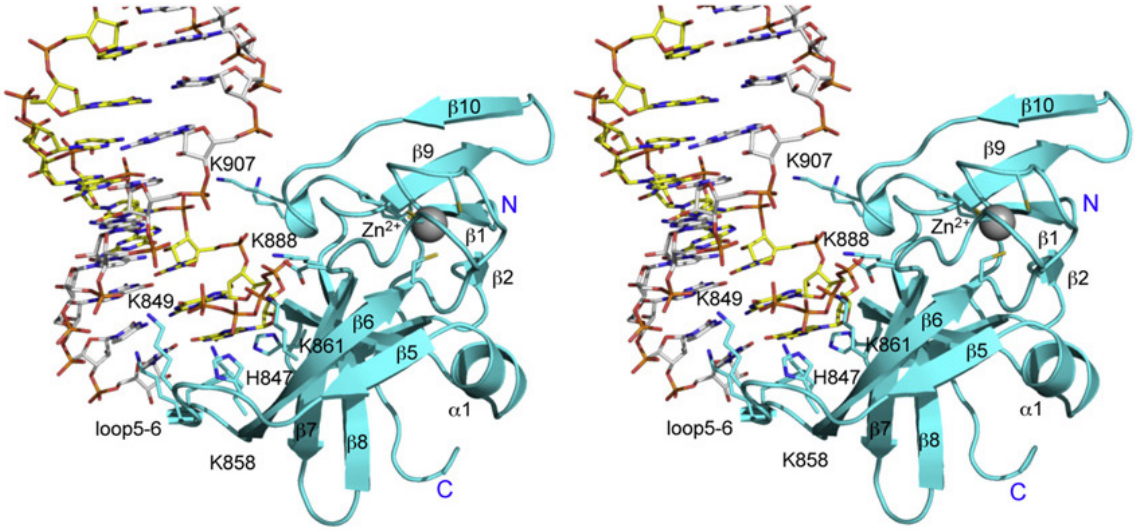


Figure 2.8 Stereo close up view of the interface between RIG-I CTD and the terminus of the 14-bp 5'-ppp dsRNA. RIG-I CTD (Chain A) is shown as cyan ribbons and the dsRNA is shown as stick models. Key residues of RIG-I CTD involved in RNA binding are shown as stick models.

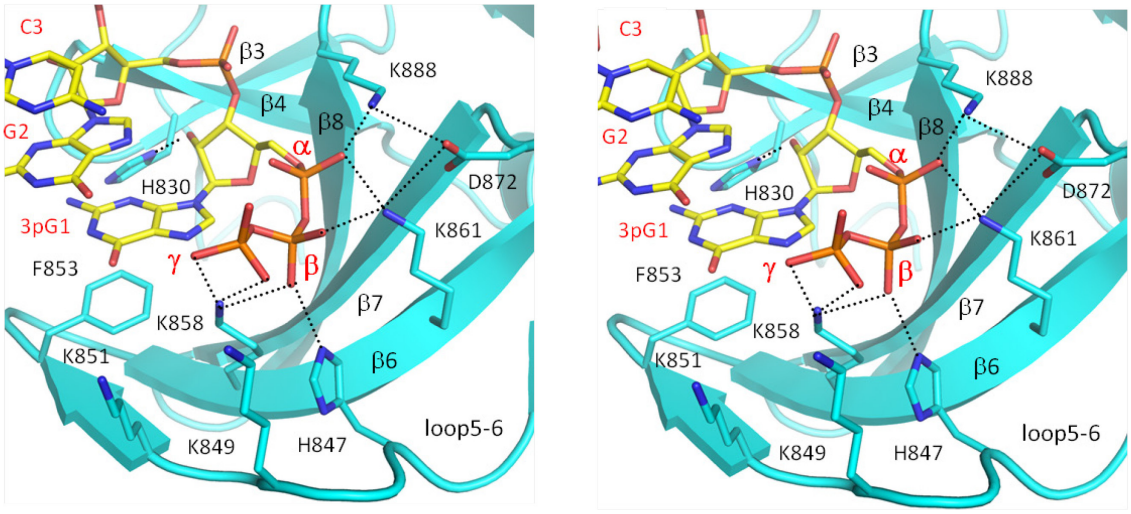


Figure 2.9 Stereo close up view of the interactions between RIG-I CTD and the 5'-ppp group of the 14-bp GC-rich dsRNA.

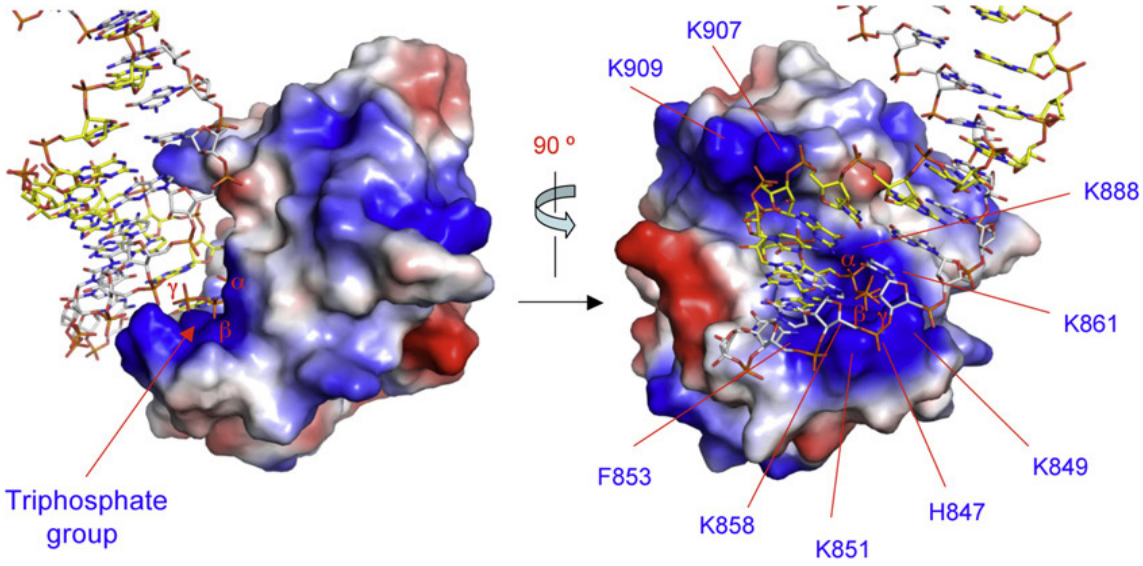


Figure 2.10 Surface representation of RIG-I CTD bound to 5'-ppp dsRNA showing the shape and charge complementarity between RIG-I CTD and the terminus of the 5'-ppp dsRNA. Positively charged surfaces are colored blue and negatively charged surfaces are red. The orientation of RIG-I CTD relative to the dsRNA on the left panel is similar to the orientation of the protein in Figure 2.8.

RIG-I CTD interacts with the dsRNA primarily through extensive electrostatic interactions with the 5'-ppp and the phosphodiester backbone (Figure 2.8 and 2.9). The exposed base pairs at the termini of the dsRNA interact with RIG-I CTD through hydrophobic interactions (Figure 2.9).

The RNA-binding surface of RIG-I CTD exhibits a high degree of shape and charge complementarity to the blunt-ends of 5'-ppp dsRNA (Figure 2.10). The total buried surface area at the RIG-CTD: dsRNA interface is approximately 1200 $Å^2$, with primary contributions from the first four nucleotides at the 5'-end of the dsRNA (buried surface area of about 900 $Å^2$). The last three nucleotides at the 3'-end of the complementary strand only make minor contributions to dsRNA binding (buried surface area of about 300 $Å^2$). The average shape correlation statistics (Sc) is 0.62 for the four dsRNA-protein interfaces; the Sc value is a measure of the degree to which two contacting surfaces are geometrically matched, with an Sc value of 1.0 indicating a perfect fit (96).

Structural basis of 5'-ppp dsRNA recognition by RIG-I CTD

As predicted from previous mutagenesis studies (52,72), the 5'-ppp binding site of RIG-I CTD is located at a large patch of positively charged surface around residues Lys858 and Lys861. Four positively charged residues interact with the 5'-ppp through electrostatic interactions (Figure 2.8 and 2.9). Lys888 interacts with the α-phosphate group through electrostatic interaction. The distance between the side chain amine group of Lys888 and the α-phosphate is only 2.4 Å. The side chain amine group of Lys861 is

about 2.7 Å away from the α- and β-phosphates, interacting simultaneously with these two phosphate groups. Proper positioning of the side chains of Lys888 and Lys861 to interact with the α- and β-phosphate is facilitated by Asp872, which interacts with the side chains of these two residues through electrostatic interactions (Figure 2.9). The β-phosphate also interacts with the side chains of Lys858 and His847 simultaneously (Figure 2.9). In addition, Lys858 also interacts with the γ-phosphate that is about 5 Å away from the side chain amine groups of Lys849 and Lys851 (Figure 2.9). These two positively charged residues may interact with the 5'-ppp through weak electrostatic interaction or solvent-mediated hydrogen bonds.

Although the electrostatic interactions between RIG-I CTD and the 5'-ppp play dominant roles in dsRNA binding, RIG-I CTD interacts extensively with other parts of the dsRNA. The exposed base pair at each terminus of the dsRNA interacts with Phe853 through hydrophobic interaction, and the guanine base at the 5'-end of the RNA stacks against the phenyl side chain of Phe853 (Figure 2.9). The side chain of His830 forms a hydrogen bond with the 2'-hydroxyl group of nucleotide G2 (Figure 2.9). This hydroxyl group also forms two solvent-mediated hydrogen bonds with the backbone amine of Tyr831 and the carbonyl group of Ile887. The phosphate group of G2 forms two solvent-mediated hydrogen bonds with Ser906 and Trp908. The third nucleotide C3 from the 5'-end forms a solvent-mediated hydrogen bond with the backbone amine of Trp908. In addition, the phosphate group of G4 interacts with the side chain of Lys907 through electrostatic interactions (Figure 2.8). However, the RNA does not interact with the nearby Lys909 directly. The side chains of Lys849 and Lys851 are in close proximity to

the phosphate of C11 and G12 near the 3'-end of the RNA (Figure 2.8). However, the electron density map for the side chains of these two residues is poorly defined, reflecting their flexibility and limited contribution to dsRNA binding. The side chain hydroxyl of Ser854 forms a hydrogen bond with the 2'-hydroxyl at the 3'-terminal nucleotide (C14) of the complementary strand. Apart from the hydrophobic interactions between the terminal bases and Phe853, no other specific interaction between RIG-I CTD and the bases was observed in the two complex structures, demonstrating that RIG-I CTD binds RNA in a sequence-independent manner.

The detailed interactions between RIG-I CTD and the 14GC 5'-ppp dsRNA are listed in Table 2.3.

RIG-I and LGP2 bind dsRNA with and without 5'-ppp in different ways

The crystal structures of RIG-I and LGP2 CTD bound to dsRNA with and without 5'-ppp, respectively, allowed us to elucidate how the RLRs recognize these two different forms of dsRNA. To facilitate this comparison, we superimposed a14-bp dsRNA onto the structure of LGP2 CTD bound to an 8-bp dsRNA (74) (Figure 2.11).

Table 2.3 Interactions between RIG-I CTD and the 14-bp GC-rich 5'-ppp dsRNA.

Nucleotide (ends) Atoms	RIG-I CTD Atoms	Distances (Å)	Type of Interactions
3pG1 (5')			
O2G	Lys858 NZ	3.4	ES
O2B	Lys858 NZ	2.7	ES
O2B	His847 NE2	2.6	ES
O1B	Lys861 NZ	2.7	ES
O2A	Lys861 NZ	2.7	ES
O2A	Lys888 NZ	2.4	ES
C2	Phe853 CE2	3.2	HP
O2'	His830 ND1	2.5	HB
O5'	Gly874 CA	3.8	VDW
O3A	Ile875 CG2	2.9	VDW
O2'	Val886 CG1	3.9	VDW
O2'	Tyr831 N (S1)	3.0 / 3.0	sHB
O2'	Ile887 O (S1)	3.0 / 3.0	sHB
G2 (5')			
O4'	His830 CE1	3.3	VDW
O2P	Lys888 CB	3.5	VDW
O2P	Trp908 N (S21)	2.6 / 3.0	sHB
O2P	Ser906 O (S32)	3.0 / 2.9	sHB
O2P	Trp908 CD1	3.2	VDW
C3 (5')			
O1P	Trp908 N (S21)	3.4/3.0	sHB
G4 (5')			
O2P	Lys907 NZ	3.2	ES
C11 (3')			
O1P	Lys849 NZ	2.7	ES (?)
C14 (3')			
O2'	Ser854 OG	3.4	HB
N3	Phe853 CB	3.9	VDW

HB: hydrogen bonds

VDW: van der Waals contacts

HP: hydrophobic interactions

sHB: solvent mediated hydrogen bonds. Solvent distances to protein atoms and to dsRNA atoms are shown.

ES: electrostatic interactions

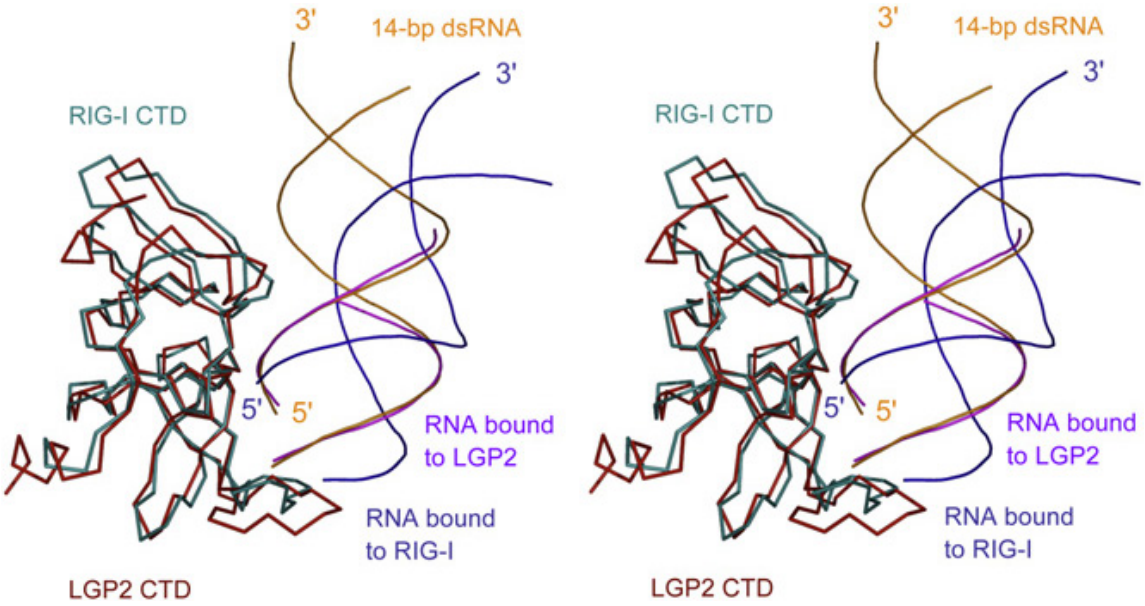


Figure 2.11 Distinct binding by RIG-I and LGP2 CTDs to dsRNA with and without 5'-ppp.

Stereo representation of the structure of RIG-I CTD bound to the 14-bp GC-rich 5'-ppp dsRNA superimposed on the structure of LGP2 CTD bound to an 8-bp dsRNA without 5'-ppp. The 14-bp 5'-ppp dsRNA bound to RIG-I CTD is shown as blue ribbons. The 8-bp dsRNA bound to LGP2 CTD is colored magenta. A 14-bp dsRNA (orange) is superimposed on the 8-bp dsRNA in the LGP2 CTD:dsRNA complex structure to show the different orientations of the dsRNA relative to the proteins in the two complexes.

Superposition of LGP2 CTD onto the structure of RIG-I CTD bound to the 14-bp GC-rich dsRNA revealed that the orientations of the dsRNA in the two complexes are dramatically different (Figure 2.11). Although both proteins recognize the terminus of the dsRNA and similar binding surfaces are involved in RNA binding, the 14-bp 5'-ppp dsRNA was rotated by approximately 30 degrees away from RIG-I CTD compared to the 14-bp dsRNA in the LGP2 CTD:dsRNA complex model (Figure 2.11). As a result, RIG-I CTD has less-extensive interactions with the 5'-ppp dsRNA compared to the interactions between the non-ppp dsRNA and LGP2 CTD (74). The total buried surface area at the LGP2 CTD:dsRNA interface (about 1,540 $Å^2$) is about 30% larger than that of the RIG-I CTD:dsRNA interface. The interface between LGP2 CTD and the non-ppp dsRNA also exhibits a higher degree of shape complementarity (Sc = 0.75).

Although the interaction between LGP2 CTD and dsRNA is more extensive, the affinity of LGP2 CTD for a 10-bp blunt-ended dsRNA without ppp was about 100 nM (74). In addition, 5'-ppp dsRNA exhibits a slightly lower affinity for LGP2 CTD (74). Furthermore, mutations that disrupted RNA binding by LGP2 do not affect its ability to regulate RIG-I signaling (74). These results suggest that LGP2 does not regulate RIG-I signaling through direct competition with RIG-I for ligand binding, indicating that the RIG-I-RNA interaction is different from LGP2-RNA interaction.

Since the 5'-ppp has to be positioned precisely into the ppp-binding site, this structural requirement determines the specific orientation of the 5'-ppp dsRNA relative to RIG-I CTD. Because the ppp makes primary contributions to 5'-ppp dsRNA binding by RIG-I CTD, the binding affinity of blunt-ended dsRNA without ppp should be much

lower than that of the 5'-ppp dsRNA; the buried surface area between a dsRNA without 5'-ppp and RIG-I CTD will be reduced to about 1050 $Å^2$ if the RNA still binds RIG-I in the same orientation. Interestingly, SPR binding studies show that RIG-I CTD binds the 14-bp dsRNA without 5'-ppp at nanomolar affinity. It is possible that RIG-I CTD binds blunt-ended dsRNA lacking a ppp in a similar way as LGP2 CTD to increase the binding interactions. Because dsRNA with and without 5'-ppp both activate RIG-I *in vivo*, we propose that the two forms of RNA may trigger the activation of RIG-I through different mechanisms.

Similar binding surfaces of RIG-I CTD mediate its interactions with different forms of RNA

Based on the structures of RIG-I CTD bound to 5'-ppp dsRNA discussed above, we expressed and purified twelve mutant proteins of RIG-I CTD and studied their binding properties with a 5'-ppp dsRNA, a non-ppp dsRNA, and a 5'-ppp ssRNA by EMSA. Ten of these twelve mutants were also analyzed for RNA binding by gel filtration chromatography.

First, mutations of each of the three positively charged lysine residues, Lys858, Lys861, and Lys888, that are involved in direct contact with the 5'-ppp to negatively charged glutamate residues almost abolished 5'-ppp dsRNA and ssRNA binding in the gel shift assays (Figure 2.12). Consistent with these results, no binding for 5'-ppp ssRNA and dsRNA were observed for these mutants by gel filtration chromatography (Figure 2.13). Similarly, mutation of Lys888 to glutamate also abolished non-ppp dsRNA

binding (Figure 2.12). However, substitution of either Lys858 or Lys861 to glutamate dramatically reduced but did not abolish non-ppp dsRNA binding (Figure 2.12 and 2.13). Substitution of His847 that interacts directly with the β-phosphate with glutamate significantly reduced the binding of 5'-ppp dsRNA and ssRNA, but only slightly reduced the binding of non-ppp dsRNA (Figure 2.12 and 2.13). Consistent with these results, mutations of both His847 and Lys861 to glutamate disrupted binding of 5'-ppp dsRNA or ssRNA and nearly abolished non-ppp dsRNA binding (Figure 2.12).

Second, we analyzed the effects of mutations of four other positively charged residues--Lys849, Lys851, Lys907, and Lys909--on RNA binding. Mutation of Lys907 that interacts with the phosphodiester backbone of the dsRNA to glutamate dramatically reduced RIG-I CTD binding for all three forms of RNA (Figure 2.12), demonstrating the critical role of this residue in recognizing the RNA backbone. Gel filtration chromatography confirmed that mutant K907E failed to bind all three forms of RNA (Figure 2.13). In contrast, substitution of the nearby Lys909 with a glutamate reduced the binding of 5'-ppp dsRNA while almost abolished the binding of non-ppp dsRNA and 5'-ppp ssRNA (Figure 2.12 and 2.13). Mutations of either Lys849 or Lys851 to glutamate reduced, but did not abolish, the binding for all three forms of RNA (Figure 2.12 and 2.13).

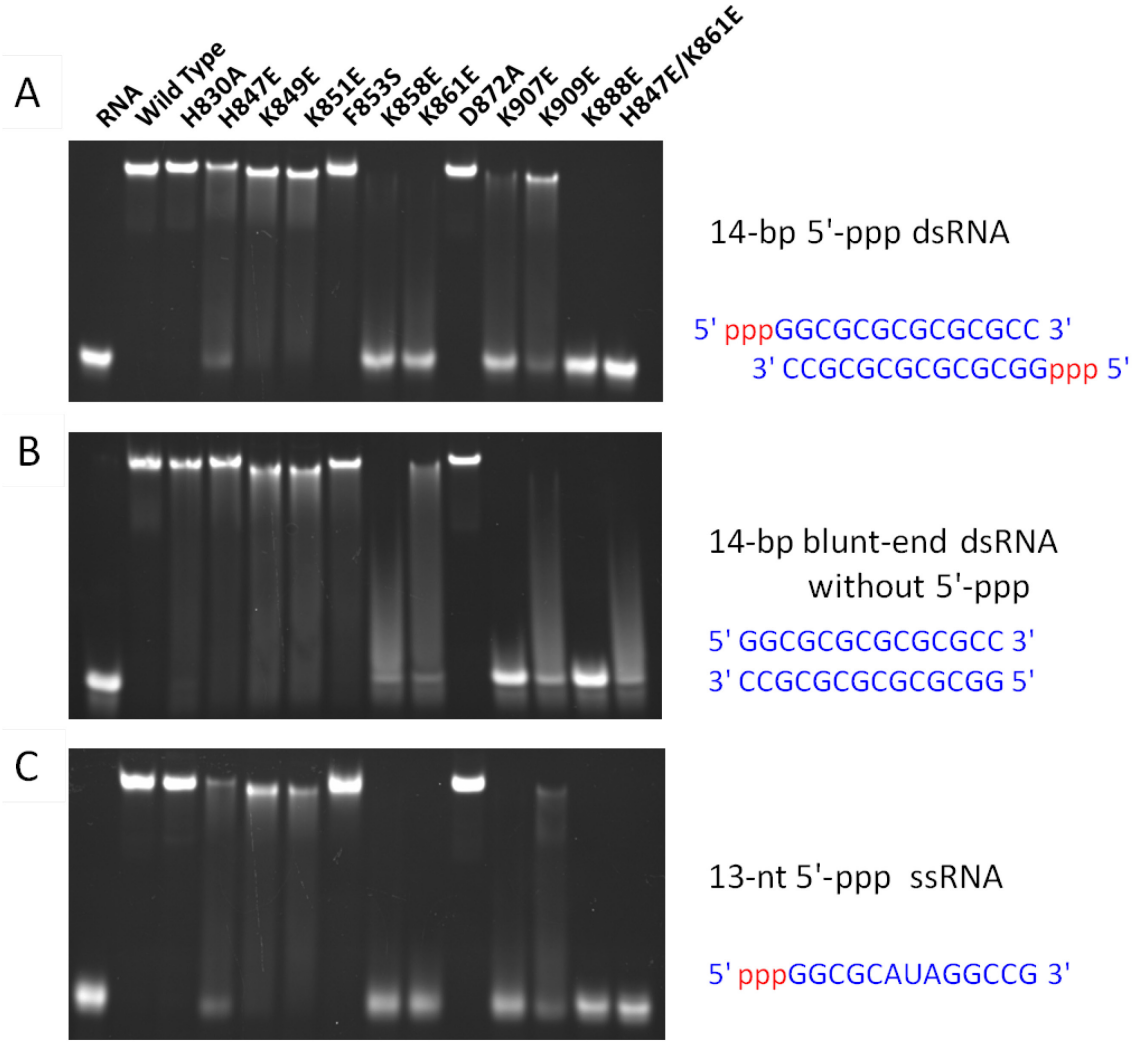


Figure 2.12 Similar binding surfaces of RIG-I CTD mediate the interactions with different forms of RNA.

(A) Binding studies of wild-type and mutants of RIG-I CTD for the 14-bp 5'-ppp dsRNA by EMSA.

(B) Binding studies of RIG-I CTD and its mutants with a 14-bp dsRNA lacking 5'-ppp.

(C) Binding studies of RIG-I CTD and its mutants with a 13-nt 5'-ppp ssRNA.

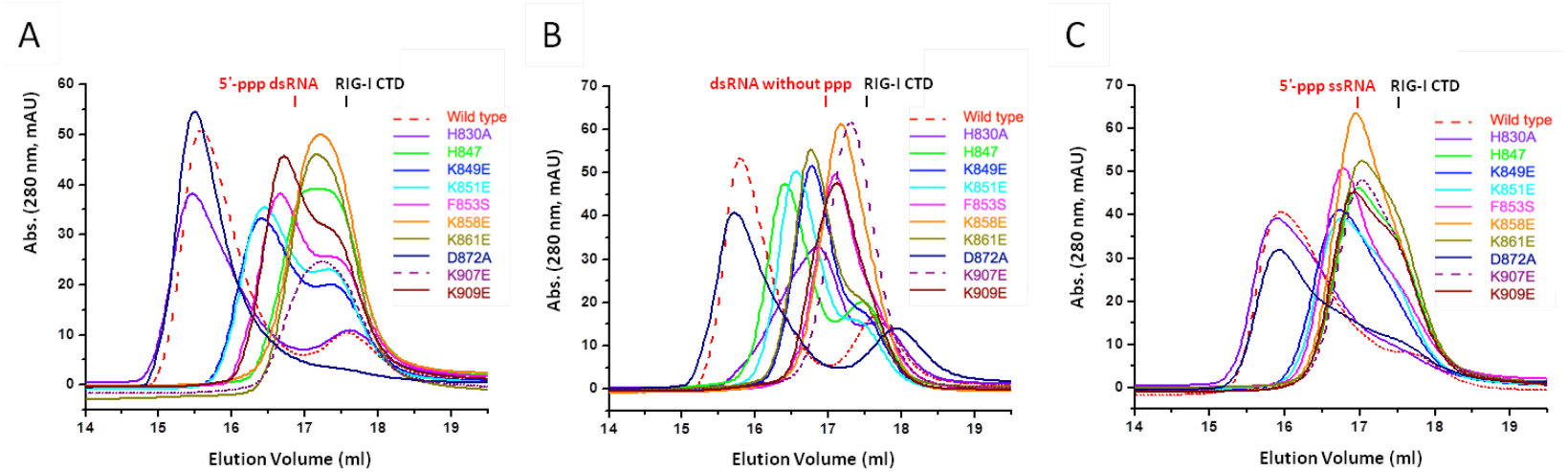


Figure 2.13 Mutations of key residues in RIG-I CTD affect the binding of three different forms of RNA.

(A) Binding studies of wild-type and mutants of RIG-I CTD for the 14-bp 5'-ppp dsRNA by gel filtration chromatography. Each of the chromatograms represents the elution profile of a 3:1 molar ratio mixture of RIG-I CTD or its mutants with the RNA. Elution volume of the protein and the RNA alone are shown above the chromatogram. Shift of the peak positions of mutant protein and RNA complexes relative to that of the wild-type protein or the RNA alone reflects how the mutation affect RNA binding. Mutants with no peak shift (such as K907E) do not bind RNA. RNA binding is not affected for mutant with peak shift (such as D872A) similar to that of the wild-type protein.

(B) Binding studies of wild-type and mutants of RIG-I CTD for the 14-bp non-ppp dsRNA.

(C) Binding studies of RIG-I CTD and its mutants for the 13-nt 5'-ppp ssRNA.

Third, we analyzed the mutations of His830, Phe853 and Asp872 that interact with the dsRNA through hydrogen bonding or hydrophobic interactions, or facilitate the binding of the ppp group. Replacement of Asp872 with an alanine did not affect RNA binding (Figure 2.12 and 2.13). Mutation of His830 to alanine did not affect 5'-ppp dsRNA or ssRNA binding, but reduced non-ppp dsRNA binding (Figure 2.13). Although the gel shift assays showed that substitution of Phe853 by serine did not affect RNA binding significantly (Figure 2.12), the gel filtration chromatography showed that this mutation dramatically reduced 5'-ppp RNA binding and nearly abolished non-ppp dsRNA binding (Figure 2.13).

These mutagenesis and binding studies demonstrated that the binding surface observed in the complex structures mediates RNA binding by RIG-I CTD in solution. It is also important to note that, although similar binding surfaces are involved in contacting non-ppp dsRNA as well as 5'-ppp dsRNA and ssRNA, the fact that mutations in the CTD can differentially affect binding of the three forms of RNA illustrates distinct recognition of these RNA ligands by RIG-I.

Mutations of key residues at the RNA-binding surface affect RIG-I signaling

The twelve mutations tested for RNA binding were built into constructs of full-length RIG-I to assess how these mutants affect RIG-I signaling in transiently transfected HEK 293T cells. The cells were first transfected to express the RIG-I mutants along with the IFN-β luciferase reporter plasmids. Expression of wild-type RIG-I and various mutants in the transfected cells was confirmed by Western blot (Figure

2.14A). After 24 hours, different RNA ligands were transfected into the cells to analyze their abilities to activate RIG-I dependent signaling.

First, we analyzed signaling of mutants modified at residues that play critical roles in the recognition of the 5'-ppp group. Because the Cα atom of Lys888 is only 3.8 Å away from the G2 phosphate and its side chain amine group interacts with the α-phosphate of the 5'-ppp, substitution of Lys888 by glutamate disrupted RNA binding by RIG-I CTD and abolished RIG-I signaling by 5'-ppp dsRNA or ssRNA and non-ppp dsRNA as well (Figure 2.14). Although mutation of Lys861 to glutamate did not affect signaling by non-ppp dsRNA, it did disrupt signaling by 5'-ppp dsRNA and ssRNA (Figure 2.14). Similarly, mutation of Lys858 to glutamate reduced signaling by non-ppp dsRNA by about 50% and abolished signaling by 5'-ppp dsRNA and ssRNA (Figure 2.14). Substitution of His847 by glutamate did not affect non-ppp dsRNA signaling, however it reduced signals from both 5'-ppp dsRNA and ssRNA dramatically (Figure 2.14).

Second, substitutions of Lys907 and Lys909 by glutamates disrupted signaling for all three forms of RNA (Figure 2.14). In contrast, mutations of either Lys851 or 849 to glutamate did not affect any of the RNA signaling (Figure 2.14).

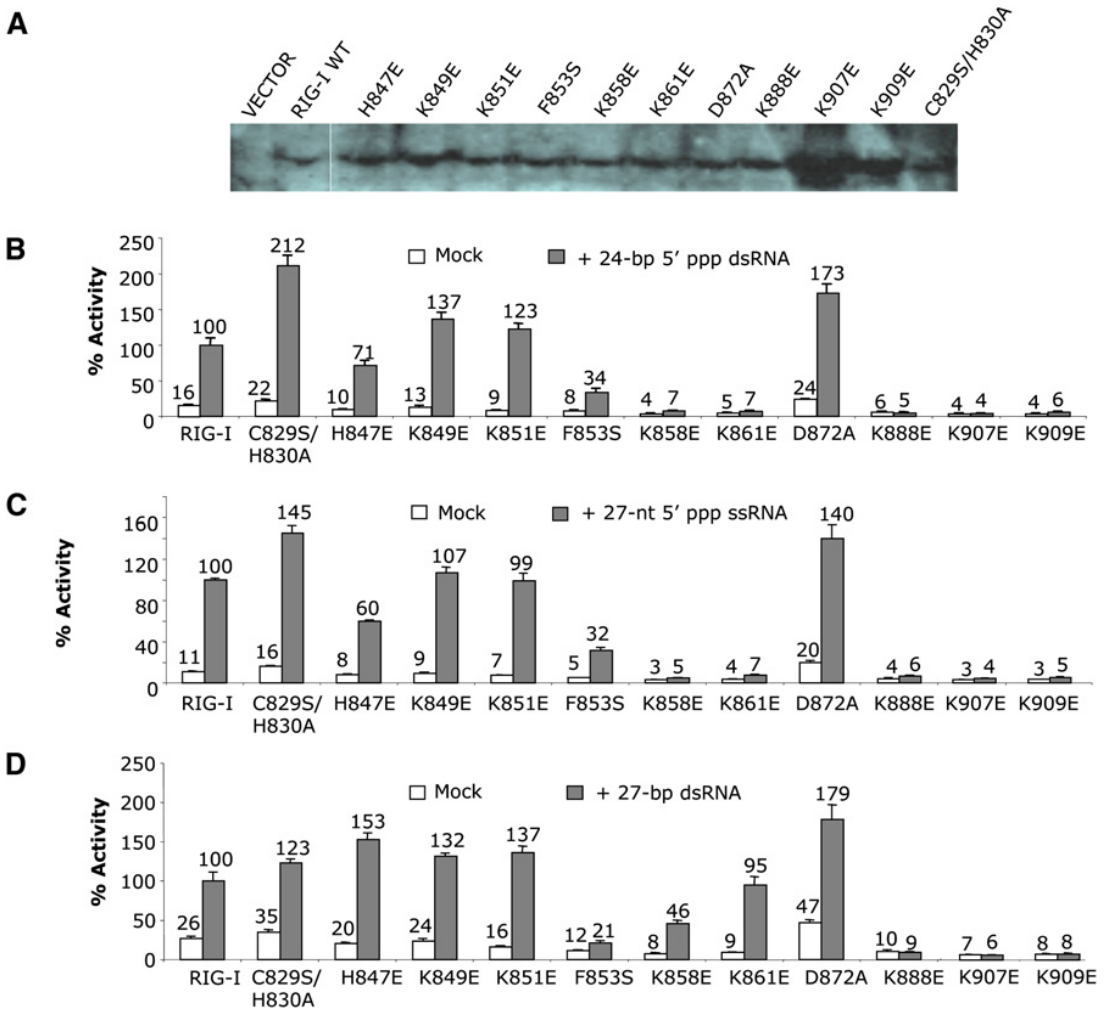


Figure 2.14 Mutations of key residues at the RNA binding surface affect RIG-I signaling.

(A) Western blot showing the expression of wild-type and mutants of full-length RIG-I in transfected cells.

(B) IFN-β luciferase assays showing the signaling of wild-type and mutants of RIG-I in HEK 293T cells stimulated with a 24-bp 5'-ppp dsRNA. Error bars correspond to the standard deviations of signals from three independent transfections.

(C) IFN-β luciferase assays showing the signaling of wild-type and mutants of RIG-I in cells stimulated with a 27-nt 5'-ppp ssRNA.

(D) IFN-β luciferase assay showing the signaling of wild-type and mutants of RIG-I in cells stimulated with a 27-bp dsRNA without 5'-ppp.

Third, we tested signaling of mutants of RIG-I at residues His830, Phe853 and Asp872 that have slightly reduced RNA binding. We found that mutations C829S/H830A or D872A had no effect on the signaling of RIG-I for any of the three forms of RNA tested (Figure 2.14). Consistent with these results, a Cys829 to serine substitution in RIG-I CTD did not affect RNA binding in vitro (Figure 2.13). The slightly higher activity of these two mutants compared to the wild-type protein is likely due to their higher expression level or better stability in the cells. However, replacement of the hydrophobic Phe853 by a hydrophilic serine reduced signaling of 5'-ppp dsRNA or ssRNA by about 70% and reduced signaling by dsRNA without 5'-ppp to background level (Figure 2.14), demonstrating that Phe853 has a role in RNA sensing by RIG-I.

These results are consistent with RNA binding properties of RIG-I CTD mutants, demonstrating that RNA binding by RIG-I CTD is essential for RIG-I signaling and RNA with and without 5'-ppp can activate RIG-I in vivo.

Discussion

RNA binding studies showed that RIG-I CTD binds 5'-ppp dsRNA or ssRNA and non-ppp dsRNA, with the highest affinity for 5'-ppp dsRNA. IFN-β luciferase assays demonstrated that all three forms of RNA could stimulate RIG-I signaling. The structures of RIG-I CTD bound to 5'-ppp dsRNA showed that RIG-I recognizes the termini of RNA and interacts with the 5'-ppp as well as the backbone phosphodiesters of the RNA. Mutagenesis and RNA binding studies demonstrated that similar binding surfaces of RIG-I CTD are involved in the binding of various forms of RNA by RIG-I.

These results provide a structural framework to understand the RNA binding properties and functions of RIG-I in antiviral immune responses.

The complex structures showed that RIG-I CTD interacts primarily with the 5' four nucleotides of the dsRNA. The complementary RNA strand only makes limited contributions to RNA binding (Figure 2.8-2.10, and Table 2.3). These observations explain how RIG-I CTD can bind both dsRNA and ssRNA with 5'-ppp.

The amino acid residues of RIG-I involved in RNA binding can be divided into three groups. The first group of residues includes Lys858, Lys861, Lys888, and His847 that recognize the 5'-ppp of dsRNA. The second group of residues includes Lys907 and Lys909 that interact with the phosphate backbone of the dsRNA. The third group of residues interacts with the exposed bases at the termini of the RNA or interacts with the backbone of the RNA by hydrogen bonds. The first two groups of residues primarily contribute to RNA binding through multiple electrostatic interactions, whereas the third group of residues makes additional contributions to RNA binding through hydrogen bonds and hydrophobic interactions. All these residues contribute to the binding of RNA with 5'-ppp, while the last two groups of residues contribute primarily to the binding of dsRNA without 5'-ppp. Importantly, these structural features explain how RIG-I can recognize such a wide range of RNA ligands including 5'-ppp dsRNA and ssRNA, as well as dsRNA without 5'-ppp.

The reason why ssRNA without 5'-ppp failed to bind RIG-I CTD is most likely due to the loss of the critical electrostatic interactions mediated by the ppp group.

However, dsRNA without ppp still binds to RIG-I CTD at high affinity. This is possibly due to the additional interactions between the complementary strand and the protein.

Because the α- and β-phosphates make major contributions to the recognition of the 5'-ppp by RIG-I (Figure 2.9), the complex structures also explain why dsRNA with one or two 5'-terminal phosphates can serve as ligands for RIG-I. Previous studies showed that blunt-ended 5'-ppp dsRNA is a potent activator of RIG-I, whereas 5'-ppp dsRNA with 5'-overhangs still stimulated RIG-I but 5'-ppp dsRNA with 3'-overhangs failed to stimulate RIG-I (64). However, our RNA binding studies showed that all three forms of dsRNA bind RIG-I CTD efficiently (Figure 2.2). Although the complex structures indicated that dsRNA with 3'-overhangs can still bind RIG-I effectively, the reason why 5'-ppp dsRNA with 3'-overhangs failed to stimulate RIG-I is most likely that these RNA molecules are unwound by RIG-I (52). The structures suggest that 5'-ppp dsRNA with 5'-overhangs should exhibit significantly reduced affinity for RIG-I just like 5'-ppp ssRNA, explaining why this kind of RNA is less effective in stimulating the activation of RIG-I (64).

Our binding studies showed that RIG-I CTD binds the GC- and AU-rich RNA with similar affinities. In addition, the two complex structures showed that RIG-I binds to the two kinds of RNA in almost the same way with no specific interactions between the protein and the bases (Figure 2.6A). These results indicate that RIG-I CTD is not responsible for the recognition of specific RNA sequences, such as the polyuridylates or polyadenylates sequences observed in the 3' untranslated regions of hepatitis C virus genome (97). The studies by Chiu *et al.* showed that poly (dA-dT) DNA transfected in

cells can activate RIG-I as a result of transcription of these DNA sequences by RNA polymerase III, whereas GC-rich DNA failed to activate RIG-I when transfected into cells (98). This is not likely due to the different sequences of DNA but most likely due to the inefficient transcription of GC rich DNA by RNA polymerase III that need the poly (dA-dT) sequence to initiate transcription. In summary, the ability of RIG-I to recognize a broad range of RNA structures without sequence specificity allows the receptor to sense the infection of a wide spectrum of viruses.

The crystal structures of RIG-I and LGP2 CTD in isolation and in complex with dsRNA are now available (60,72,74), as well as the solution structure of MDA5 CTD (73). These structures revealed that the CTD of the RLRs are highly conserved and similar binding surfaces are involved in RNA binding by all the three proteins (Figure 2.7). The structures of RIG-I CTD and LGP2 CTD bound to dsRNA with and without 5'-ppp demonstrated that the RLRs recognize the termini rather than the ribose-phosphate backbone of various forms of RNA (Figure 2.11). However, the three proteins exhibit different ligand-binding properties and play different roles in antiviral immune responses. Comparisons of the structures of RIG-I CTD in complex with 5'-ppp dsRNA and LGP2 CTD in complex with non-ppp dsRNA revealed that the orientations of the RNA are significantly different (Figure 2.11). This finding suggests that the 5'-ppp and non-ppp dsRNA interact with RIG-I differently and probably induce different conformations of full-length RIG-I upon binding. This idea is further supported by the differential binding of 5'-ppp dsRNA and non-ppp dsRNA by RIG-I CTD mutants

(Figure 2.12 and 2.13) and the ability of RIG-I mutants to activate luciferase reporters in cells stimulated with different forms of RNA (Figure 2.14).

Comparisons of the ligand-binding surfaces of the RLR CTDs showed that the surface electrostatic potential of the three proteins are dramatically different (Figure 2.15). For the four positively-charged residues in RIG-I CTD that are responsible for the 5'-ppp recognition, only Lys888 is conserved in the sequences of LGP2 and MDA5 CTDs (Figure 1.2). Consistent with this observation, both MDA5 and LGP2 CTDs bind 5'-ppp dsRNA with lower affinities compared to blunt-ended dsRNA without 5'-ppp (74). It is evident that RIG-I is the only RLR that is responsible for the sensing of 5'-ppp RNA. Because RIG-I can also be activated by dsRNA without 5'-ppp, which is also recognized by MDA5, this could explain why RIG-I and MDA5 sense distinct but overlapping sets of viruses.

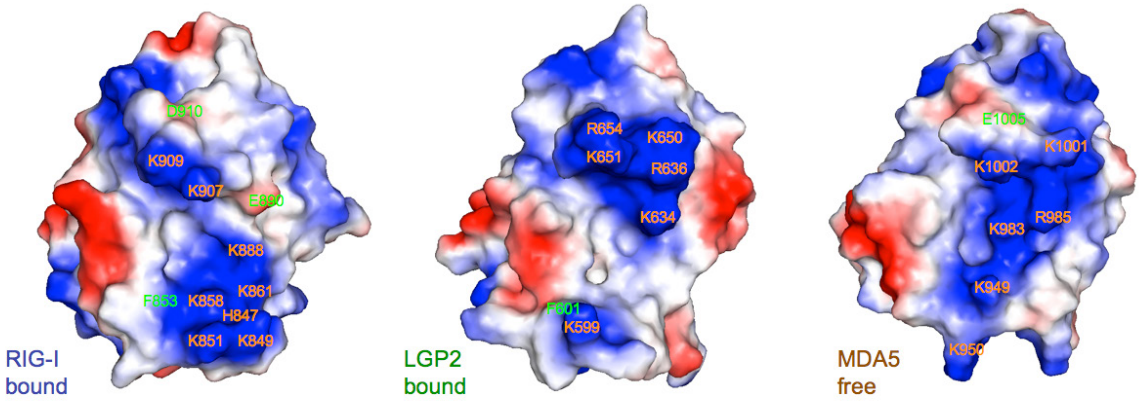


Figure 2.15 Electrostatic surface potential of the RNA binding surfaces of RIG-I, LGP2, and MDA5 CTDs.
Positively charged regions are colored blue, while negatively charged regions are red.

A number of factors may affect the results of RIG-I activation assays *in vivo*. The structure and purity of RNA ligands, their binding affinities for RIG-I, their effective concentrations in cells, and their stabilities will all affect the signal readout. Previous studies used various cell lines and different assays to demonstrate that various RNA ligands can activate signal transduction through RIG-I (90). Our studies demonstrated that dsRNA with and without 5'-ppp, and 5'-ppp ssRNA are the three major families of RNA that activate RIG-I. Furthermore, blunt-ended dsRNA with and without 5'-ppp are more effective agonists for RIG-I signaling than dsRNA with overhangs (52,64,74,90). Consistent with these findings, all these three families of RNA bind to RIG-I CTD, and blunt-ended dsRNA with 5'-ppp has a significantly higher affinity (Figure 2.1 and 2.4). Results from the IFN-β luciferase reporter assays confirmed that all these three families of RNA stimulate RIG-I signaling in cells (Figure 2.14). In contrast, two recent studies showed that only dsRNA with 5'-ppp activate RIG-I in human monocytes (64,65). It is likely that the different assay techniques or the difference in the effective concentration of ligands in the transfected cells account for these discrepancies. Because the affinities of blunt-ended dsRNA without ppp and 5'-ppp ssRNA are about 15-fold lower than that of blunt-ended 5'-ppp dsRNA, significantly higher concentration of non-ppp dsRNA is needed for the activation of RIG-I. A number of other reports, including one of our previous studies, also demonstrated that dsRNA without 5'-ppp can activate RIG-I (52,74). In addition, it has been shown that RNase digestion of dsRNA mimetic poly I:C converts it from a ligand of MDA5 into a ligand for RIG-I (71). Moreover, the digestion of self-RNA by antiviral endonuclease RNase L generates small RNA products that

initiate interferon production (30). It is likely that higher concentration of ligands generated by RNase digestion triggers the activation of RIG-I.

Structural studies of RIG-I and LGP2 CTD bound to dsRNA provided important insights into the structural basis of viral RNA recognition by the RLRs. However, the mechanism of how RNA binding triggers the activation of RIG-I or MDA5 remains to be established. Although electron microscopy studies of RIG-I and LGP2 in isolation and in complex with dsRNA provided some information about how these proteins might bind RNA (99,100), the detailed mechanism of how RNA binding activates the RLRs awaits the determination of high resolution structures of the full-length proteins bound to RNA.

CHAPTER III

STRUCTURAL BASIS OF THE INTERACTIONS BETWEEN BLUNT-ENDED DOUBLE-STRANDED RNA WITHOUT 5'-TRIPHOSPHATE AND RIG-I C-TERMINAL DOMAIN*

Background

RIG-I recognizes molecular patterns in viral RNA to regulate the induction of type I IFNs. Since RNA can fold into complicated three-dimensional structures and contain various modifications, especially at their termini, it is not clear what the exact structural feature of the viral RNA is recognized by RIG-I and how RNA binding activates RIG-I signaling.

RNA transcripts containing a 5'-ppp potently trigger the activation of RIG-I upon transfection into cells (61,62). Recent studies using chemically synthesized dsRNA or hairpin RNA containing 5'-ppp further demonstrated that base pairing near the 5'-ppp is necessary for RIG-I activation (64,65,101). Moreover, blunt-ended dsRNA without 5'-ppp can also activate RIG-I signaling in transfected cells (66,74). In addition, digestion

---

of the dsRNA mimetic poly I:C by RNase III to fragments of <1 kb transforms poly I:C from an activator of MDA5 to an activator of RIG-I (71). Consistent with this, short RNA generated by RNase L digestion of cellular RNA induced IFN production via RIG-I and MDA5 (30). These results indicated that the 5'-ppp is not absolutely needed for the activation of RIG-I by RNA. Clearly, how RIG-I recognizes such a wide range of ligands need to be better defined.

Studies in Chapter II showed that the CTD of RIG-I exhibits high affinity for 5'-ppp dsRNA as well as non-ppp dsRNA. The crystal structure of LGP2 CTD bound to an 8-bp blunt-ended non-ppp dsRNA provided the first evidence that the termini of dsRNA are recognized by the RLRs (74). Structures of RIG-I CTD bound to the 14-bp GC-rich and the 12-bp AU-rich 5'-ppp dsRNA confirmed that RIG-I CTD also recognizes the termini of 5'-ppp dsRNA, and interacts with the ppp group through electrostatic interactions. A similar mode of RNA binding was observed in the structure of RIG-I CTD bound to a different 12-bp dsRNA (102). These structures showed that the 5'-ppp of the dsRNA is recognized by a cluster of positively charged residues including Lys858, Lys861, Lys888, and His847. In contrast, both MDA5 and LGP2 CTDs bind to blunt-ended dsRNA with no preference for 5'-ppp group. Consistent with this, structures of MDA5 and LGP2 CTDs showed that these two proteins lack the ppp-binding site observed in RIG-I CTD. Comparisons of the structures of RIG-I CTD bound to 5'-ppp dsRNA and LGP2 CTD bound to non-ppp dsRNA showed that the dsRNA were in dramatically different orientations. These observations led us to hypothesize that RIG-I

CTD will bind non-ppp dsRNA and 5'-ppp dsRNA in different ways. However, the structural basis for the recognition of non-ppp dsRNA by RIG-I is not fully understood.

To understand to promiscuous RNA binding by RIG-I, we determined the crystal structure of RIG-I CTD bound to a 14-bp non-ppp dsRNA at 2.4-Å resolution and compared the structure of RIG-I CTD bound to 5'-ppp dsRNA of the same sequence. These two structures revealed that distinct but overlapping sets of residues are involved in the binding of dsRNA with and without 5'-ppp. The orientation of non-ppp dsRNA relative to RIG-I CTD is dramatically different from that of the 5'-ppp dsRNA. Mutagenesis of key residues at the RNA-binding surface affected RNA binding *in vitro* and RIG-I signaling *in vivo*.

Materials and methods

RNA binding studies

Human RIG-I CTD was expressed and purified as described in Chapter II. Different forms of RNA were generated either by chemical synthesis (IDT) or by *in vitro* transcription using T7 RNA polymerase as described in Chapter II. The *in vitro* transcribed RNA were purified by gel filtration chromatography and analyzed by denaturing PAGE. Sequences of all the RNA and DNA oligos used in this study are shown in Table 3.1. All the dsRNA used were generated by annealing of the ssRNA, as described in Chapter II.

Table 3.1 Sequences of RNA and DNA used in this study.

The 5'-ppp RNA were synthesized by *in vitro* transcription and purified by gel filtration chromatography or denaturing PAGE. The dsRNA without 5'-ppp were chemically synthesized by IDT.

14-bp GC-rich 5'-ppp dsRNA	5' pppGGCGCGCGCGCC 3' 3' CCGCGCGCGCGGppp 5'
14-bp blunt-ended dsRNA without 5'-ppp	5' GGCGCGCGCGCC 3' 3' CCGCGCGCGCGG 5'
14-bp blunt-ended dsDNA	5' GGCGCGCGCGCC 3' 3' CCGCGCGCGCGG 5'
13-nt 5'-ppp ssRNA	5' pppGGCGCAUAGGCCG 3'
13-bp RNA:DNA hybrid	5' GGCGCAUAGGCCG 3' (RNA) 3' CCGCGTATCCGGC 5' (DNA)
14-bp dsRNA with 5'-overhangs	5' AUAUGGCGCGCGCGCC 3' 3' CCGCGCGCGCGGUAUA 5'
14-bp dsRNA with 3'-overhangs	5' GGCGCGCGCGCCAUAU 3' 3' UAUACCGCGCGCGCGG 5'
24-bp 5'-ppp dsRNA	5' pppGGCGCGCGAGUCGACUCGCGCGCC 3' 3' CCGCGCGCUCAGCUGAGCGCGCGGppp 5'
27-nt 5'-ppp ssRNA	5' pppGGUGCAGAUGAACUUCAGGGUCAGCUU 3'
27-bp blunt-ended dsRNA without 5'-ppp	5' AAGCUGACCCUGAAGUUCAUCUGCACC 3' 3' UUCGACUGGGACUUCAAGUAGACGUGG 5'

In the binding studies by gel filtration chromatography, the RNA samples (at about 100 μM) were mixed with equal volume of RIG-I CTD (at about 200 μM) and 100 μl of samples were injected over a Superdex200 (10/300 GL) column (GE Healthcare) eluted with the Tris-NaCl buffer.

In the binding studies by EMSA, each type of RNA was mixed with RIG-I CTD at a molar ratio 1:3 at a final concentration of about 1.6:5 μM. The mixtures were resolved on 12% native polyacrylamide gels and stained with EtBr.

Purification, crystallization and structural determination of RIG-I CTD and 14-bp dsRNA complex

Human RIG-I CTD was expressed and purified as described in Chapter II. The 14-bp blunt-ended non-ppp dsRNA (Table 3.1) was chemically synthesized by IDT. The dsRNA was mixed with excess RIG-I CTD and the 2:1 (RIG-I CTD:dsRNA) complex was purified by gel filtration chromatography on a Superdex75 (1.6 × 60) column (GE Healthcare).

The complex was concentrated to about 10 mg/ml and crystallized in 27% PEG 500 monomethyl ether (PEG MME 500) in 0.1 M HEPES buffer at pH 7.5. The crystals were flash frozen in liquid nitrogen.

Diffraction data were collected using a Rigaku RAXIS IV++ detector and processed with HKL2000 (92). The complex crystallizes in space group $P2_12_12_1$ with cell dimensions: a = 36.99 Å, b = 70.30 Å, c = 125.54 Å. The crystallographic ASU contains one 2:1 (RIG-I CTD:dsRNA) complex. Structure of RIG-I CTD in the complex

was determined by molecular replacement with MOLREP (94) using the structure of RIG-I CTD bound to the 14-bp 5'-ppp dsRNA (as described in Chapter II) as search model (PDB code: 3LRN). The structural model was rebuilt using O (95). After several rounds of refinement with CNS (93), electron density for the dsRNA became obvious. A 14-bp dsRNA was manually docked into the electron density map and rebuilt with O. The structure was refined using CNS.

The atomic coordinates and structure factors of RIG-I CTD bound to the 14-bp non-ppp dsRNA have been deposited with the RCSB Protein Data Bank under the accession code 3OG8. The structural figures were generated with PyMol (http:///www.pymol.org).

Mutagenesis studies of dsRNA binding surface of RIG-I CTD

For dsRNA binding surface analysis, 4 residues were mutated for binding studies. Point mutations were introduced by PCR using a QuikChange Site-Directed Mutagenesis kit (Stratagene). DNA primers (IDT) used in this study were:

For R811E

Forward primer

5'- GGAAAATAAAAAACTGCTCTGCGAGAAGTGCAAAGCCTTGGCATGTT -3'

Reverse primer

5'- AACATGCCAAGGCTTTGCACTTCTCGCAGAGCAGTTTTTTATTTTCC -3'

For R811S

Forward primer 5'- AAAAAACTGCTCTGCAGCAAGTGCAAAGCCTTGGC -3'

Reverse primer 5'- GCCAAGGCTTTGCACTTGCTGCAGAGCAGTTTTTT -3'

For R812E

Forward primer

5'- AAAATAAAAAACTGCTCTGCAGAGAGTGCAAAGCCTTGGC -3'

Reverse primer 5'- GCCAAGGCTTTGCACTCTCTGCAGAGCAGTTTTTTATTTT -3'

For R812S

Forward primer

5'- AAAAAACTGCTCTGCAGAAGCTGCAAAGCCTTGGCATGT -3'

Reverse primer 5'- ACATGCCAAGGCTTTGCAGCTTCTGCAGAGCAGTTTTTT -3'

For R866E

Forward primer

5'- GAGCAAAGATATTCTGTGCCGAGCAGAACTGCAGCCATGACTG -3'

Reverse primer

5'- CAGTCATGGCTGCAGTTCTGCTCGGCACAGAATATCTTTGCTC -3'

For H871E

Forward primer 5'- CCGACAGAACTGCAGCGAGGACTGGGGAATCCATG -3'

Reverse primer 5'- CATGGATTCCCCAGTCCTCGCTGCAGTTCTGTCGG -3'

Sequences of the mutants were confirmed by plasmid DNA sequencing (Institute of plant genomics and biotechnology, Texas A&M University).

The mutant proteins were expressed and purified the same way as the wild-type protein. RNA binding studies for the mutants were conducted by EMSA and gel filtration chromatography.

Mutagenesis of RIG-I and IFN-β luciferase reporter gene assays

IFN-β luciferase reporter gene assays were used to analyze the effects of mutations at RNA-binding surface on RIG-I signaling in HEK 293T cells, with the helps from our collaborators at the Indiana University, Bloomington (Dr. C. Cheng Kao group, Department of Molecular and Cellular Biochemistry), according to the methods described in Chapter II. Briefly, the cells were transfected with plasmid of wild-type RIG-I or its mutans along with IFN-β luciferase reporter plasmid and Renilla luciferase control plasmid. After 24 hr, three different forms of RNA (Table 3.1) at 50 nM concentrations were transfected into the cells to stimulate RIG-I dependent signaling. The ratios of firefly luciferase reporter over Renilla luciferase control were determined.

Mutations of full-length RIG-I were performed using pUNO-hRIG-I (Invivogen) as template. Expression of wild-type and mutants of RIG-I in the transfected cells were confirmed by western blot using anti-RIG-I antibody (C-15, sc-48929, Santa Cruz biotechnology).

Results

RIG-I CTD binds to blunt-ended dsRNA without 5'-ppp

We chemically synthesized a 14-bp blunt-ended dsRNA, two 14-bp dsRNA with either 5'- or 3'-overhangs, a 14-bp dsDNA, and a 13-bp RNA:DNA hybrid (Table 3.1), and studied their binding interactions with RIG-I CTD by both gel filtration chromatography and EMSA.

As discussed in Chapter II, gel filtration chromatography showed that RIG-I CTD binds to the blunt-ended non-ppp dsRNA and forms a stable complex (Figure 2.1D and 3.1A). Based on the gel filtration chromatograms of the RIG-I CTD:dsRNA mixtures at different molar ratios (Figure 3.1A), the stoichiometry between RIG-I CTD and the 14-bp blunt-ended dsRNA is estimated to be 2:1. Binding studies by EMSA also showed that RIG-I CTD binds to blunt-ended non-ppp dsRNA as well as 5'-ppp dsRNA (Figure 3.1F).

However, the 14-bp non-ppp dsRNA with 5'-overhangs failed to bind RIG-I CTD (Figure 3.1B and F). Although the dsRNA with 3'-overhangs still binds to RIG-I CTD by EMSA (Figure 3.1F), gel filtration chromatography showed that it binds to RIG-I CTD with reduced affinity and does not form the stoichiometric 2:1 complex (Figure 3.1C).

In addition, gel filtration chromatography showed that RIG-I CTD does not bind a 14-bp dsDNA with the same sequence as the 14-bp blunt-ended dsRNA (Figure 3.1D). Furthermore, a 13-bp blunt-ended RNA:DNA hybrid also binds to RIG-I CTD (Figure 3.1E). The stoichiometry between RIG-I CTD and the RNA:DNA hybrid is estimated to be 1:1.

Figure 3.1 RIG-I CTD binds to blunt-ended dsRNA without 5'-ppp.
(A) Binding studies of RIG-I CTD with the 14-bp blunt-ended dsRNA by gel filtration chromatography. RIG-I CTD, the 14-bp blunt-ended dsRNA, and mixtures of RIG-I CTD:dsRNA at molar ratios 1:1 to 4:1 were injected over a superdex200 column. Elution profile of RIG-I CTD is in black and the dsRNA in red. Elution profiles of mixtures of RIG-I CTD and the 14-bp dsRNA at molar ratios 1:1, 2:1, 3:1, and 4:1 are in green, blue, cyan, and purple, respectively. The molecular masses of four protein standards and their elution positions are shown above the chromatograms.
(B) Binding studies of RIG-I CTD with a 14-bp dsRNA containing 5'-AUAU overhangs.
(C) Binding studies of RIG-I CTD with a 14-bp dsRNA containing 3'-AUAU overhangs.
(D) Binding studies of RIG-I CTD with a 14-bp blunt-ended dsDNA with the same sequence as the 14-bp blunt-ended dsRNA.
(E) Binding studies of RIG-I CTD with a 13-bp RNA:DNA hybrid.
(F) Binding studies of RIG-I CTD with the 14-bp 5'-ppp dsRNA, 14-bp blunt-ended non-ppp dsRNA, and 14-bp non-ppp dsRNA with either 5'- or 3'-overhangs by EMSA.

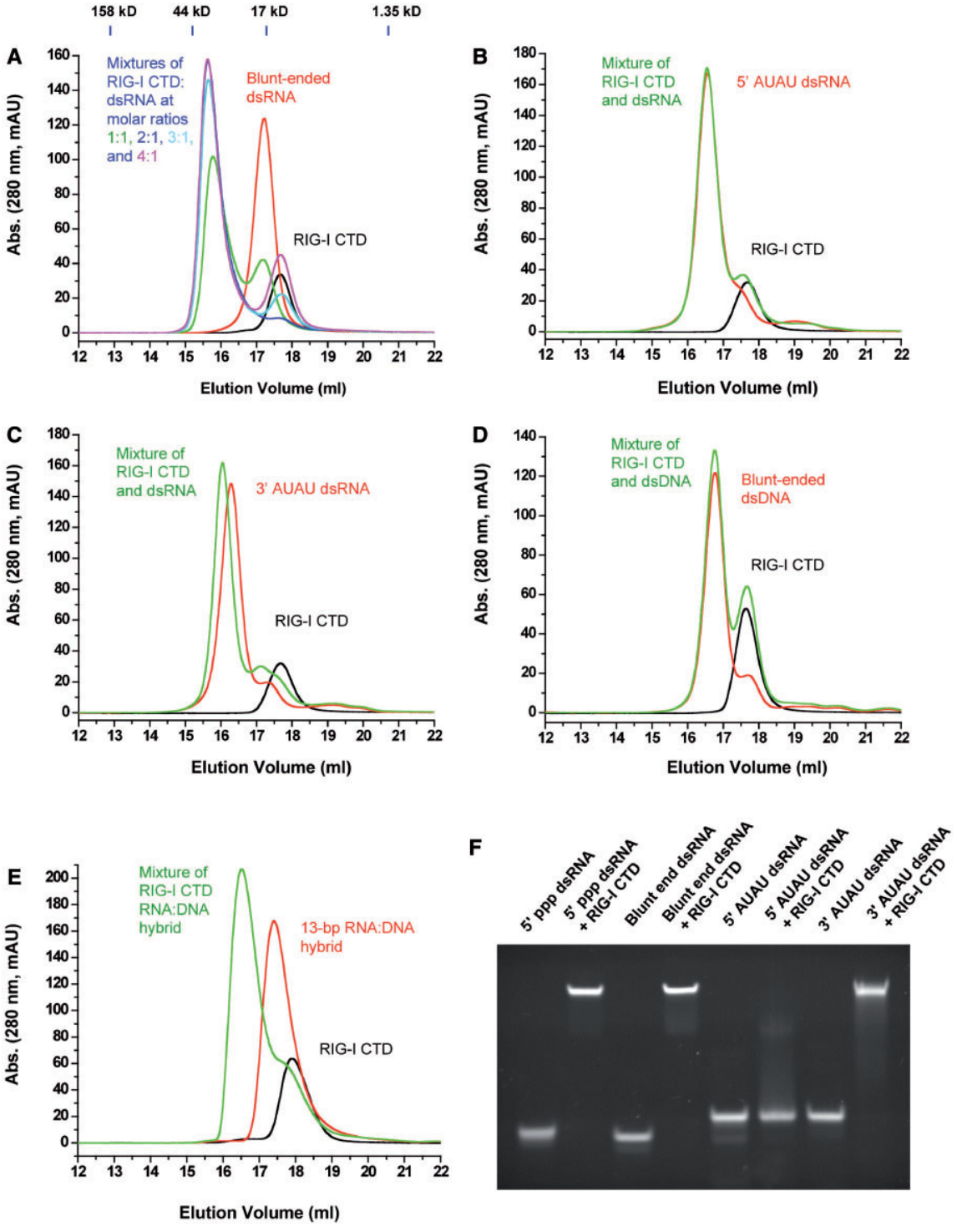

158 kD
44 kD
17 kD
1.35 kD
A
Mixtures of RIG-I CTD: dsRNA at molar ratios 1:1, 2:1, 3:1, and 4:1
Blunt-ended dsRNA
RIG-I CTD
Abs. (280 nm, mAU)
Elution Volume (ml)
B
Mixture of RIG-I CTD and dsRNA
5' AUAU dsRNA
RIG-I CTD
C
Mixture of RIG-I CTD and dsRNA
3' AUAU dsRNA
RIG-I CTD
D
Mixture of RIG-I CTD and dsDNA
Blunt-ended dsDNA
RIG-I CTD
E
Mixture of RIG-I CTD RNA:DNA hybrid
13-bp RNA:DNA hybrid
RIG-I CTD
F
5' ppp dsRNA
5' ppp dsRNA + RIG-I CTD
Blunt end dsRNA
Blunt end dsRNA + RIG-I CTD
5' AUAU dsRNA
5' AUAU dsRNA + RIG-I CTD
3' AUAU dsRNA
3' AUAU dsRNA + RIG-I CTD

Table 3.2 Statistics of crystallographic analysis.

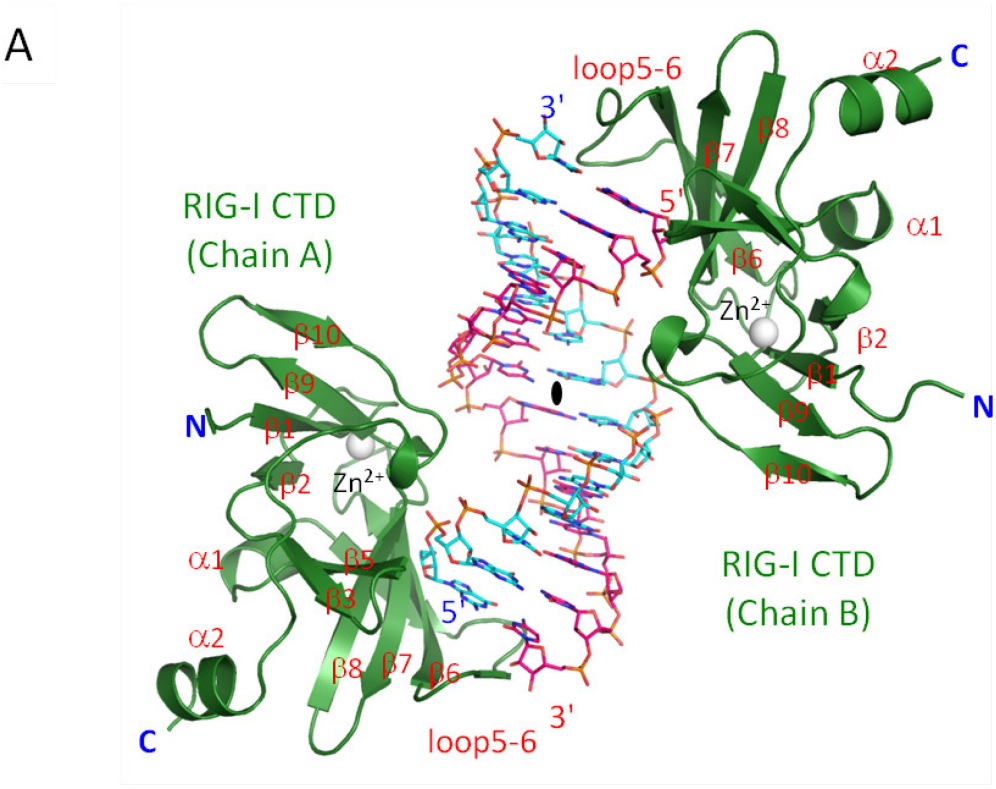
Diffraction Data	
Space group	$P2_12_12_1$
Unit cell	a = 36.99 Å, b = 70.30 Å, c = 125.54 Å
Asymmetric unit	One 2:1 RIG-I CTD:dsRNA complex
Resolution (Å)	50.0-2.40 (2.49-2.40)[a]
Unique reflections	13384
Redundancy	6.0 (5.8)
Completeness (%)	99.1 (98.0)
$I/\sigma I$	29.5 (5.3)
R_{sym} (%)	8.6 (44.8)
Refinement	
Resolution (Å)	50.0-2.40
Reflections (F > 0) (total/test set)	12860/1329
No. of atoms Protein RNA Zinc Ion Solvent	 20998 596 2 133
R_{cryst}/R_{free} (%)	22.1/28.5
R.m.s.d. bond length R.m.s.d. bond angle	0.006 Å 1.32°

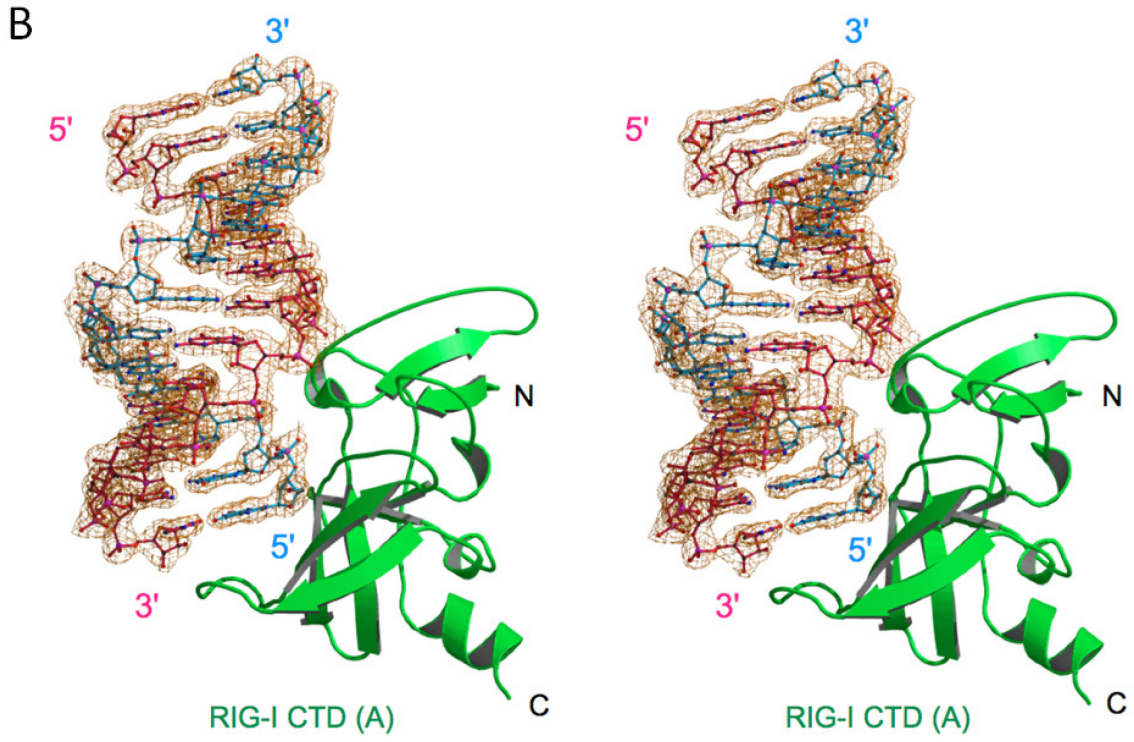
[a] Values in the parentheses are for the highest-resolution shell; 10% of reflections are used in the test set for R_{free} calculation.

Figure 3.2 Crystal structure of human RIG-I CTD bound to a 14-bp blunt-ended dsRNA without 5'-ppp (PDB code 3OG8).

(A) Structure of RIG-I CTD bound to the dsRNA. RIG-I CTDs are shown by the green ribbons. The dsRNA is shown by the sticks representation. Carbon atoms of the two RNA strands are colored cyan and pink, respectively. The zinc ion bound to RIG-I CTD is shown by the gray sphere. The complex exhibits pseudo two-fold non-crystallographic symmetry. The pseudo two-fold axis is shown by the black oval.

(B) Stereo representation of the σ_A weighted $2F_o$ - F_c map (orange mesh, contoured at 1.0σ) for the 14-bp blunt-ended non-ppp dsRNA bound to RIG-I CTD. The dsRNA is shown as pink and cyan ball-and-stick models. RIG-I CTD is shown as green ribbons.


A
RIG-I CTD
(Chain A)
loop5-6
α2
C
3'
β8
β7
5'
α1
β6
Zn2+
β2
β10
β9
β1
N
N
β1
β9
β2
Zn2+
β10
α1
β5
β3
5'
RIG-I CTD
(Chain B)
α2
β8
β7
β6
3'
C
loop5-6


B
3'
5'
N
5'
3'
RIG-I CTD (A)
3'
5'
N
5'
3'
RIG-I CTD (A)

Crystal structure of RIG-I CTD bound to blunt-ended non-ppp dsRNA

To elucidate the structural basis of non-ppp dsRNA recognition by RIG-I, we cocrystallized RIG-I CTD (residues 803 to 923) with a 14-bp blunt-ended dsRNA. The crystal structure of the complex was determined by molecular replacement and refined at 2.40-Å resolution (PDB code 3OG8). Statistics of data collection and structural refinement are shown in Table 3.2.

The overall structure of the complex crystal is shown in Figure 3.2. Each crystallographic ASU contains a 2:1 complex between RIG-I CTD and the dsRNA. The structure showed that one RIG-I CTD bound to each terminus of the blunt-ended dsRNA, forming a complex with pseudo two-fold symmetry (Figure 3.2A). Structures of the two RIG-I CTDs in the complex are essentially the same (r.m.s.d 0.6 Å) and the two RIG-I CTDs bind to the dsRNA in similar ways. Discussions of the structure will be based on the interface between RIG-I CTD chain A and the dsRNA due to better-defined electron density throughout this interface (Figure 3.2B).

The overall structure of the RIG-I CTD:dsRNA complex looks like two hands holding a stack of chips (Figure 3.2A). The exposed base pairs at the termini of the dsRNA stack on the loop connecting strands β5 and β6 (loop 5-6), and interact with RIG-I CTD through hydrophobic interactions. The thumb, corresponding to the loop connecting strand β10 and the C-terminal helix, reaches into the major groove of the dsRNA and interacts with the RNA through electrostatic interactions.

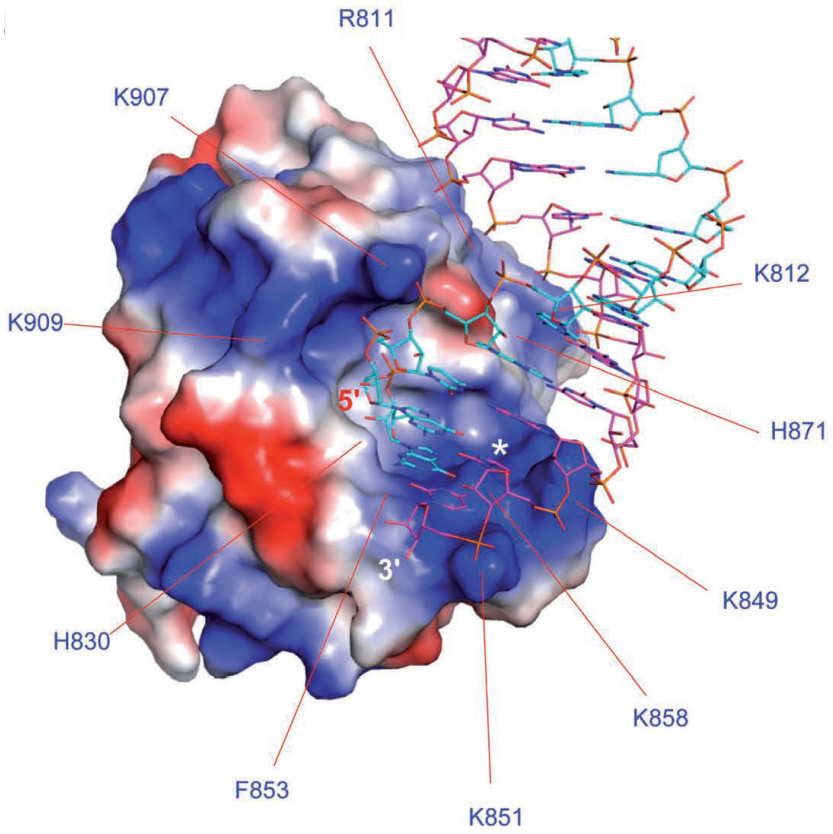


Figure 3.3 Surface electrostatics of RIG-I CTD.

Positively charged surface is colored blue and negatively charged surface red. The blunt-ended non-ppp dsRNA bound to RIG-I CTD is shown by the stick models. Key residues mediating non-ppp dsRNA recognition are labeled. Location of the ppp-binding site for 5'-ppp dsRNA is indicated by the white asterisk.

Figure 3.4 Structural basis of blunt-ended non-ppp dsRNA recognition by RIG-I CTD.
(A) Interactions between the 5' four nucleotides of the 14-bp blunt-ended non-ppp dsRNA and RIG-I CTD. The dsRNA is shown by the stick models. Carbon atoms of the RNA strand interacting with RIG-I CTD are colored cyan. The complementary strand is colored light gray. RIG-I CTD is shown by the green ribbons. Key residues involved in RNA binding are shown by the magenta stick models.
(B) Interactions between complementary strand and RIG-I CTD. The complementary strand of the RNA is shown by the stick models with carbon atoms colored pink.
(C) Schematic representations of the interactions between RIG-I CTD and the 14-bp non-ppp dsRNA (left) or the 14-bp 5'-ppp dsRNA (right). Interactions between the dsRNA and one of the two RIG-I CTD molecules are shown.

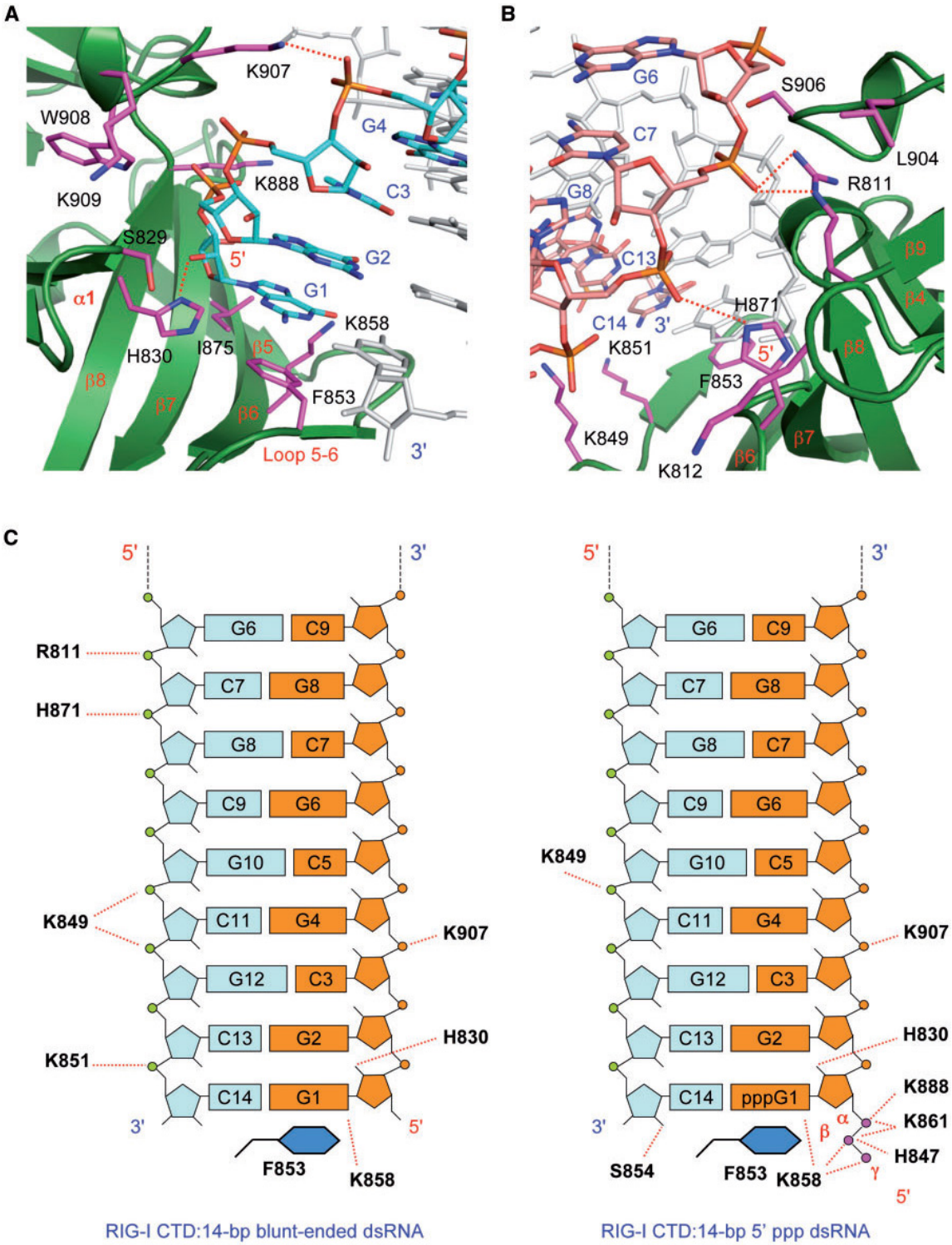

A
K907
W908
G4
K888
C3
K909
S829
5'
G2
G1
α1
K858
H830
I875
β5
β8
β7
β6
F853
Loop 5-6
3'
B
G6
S906
C7
L904
R811
G8
C13
β9
β4
H871
C14
3'
K851
5'
β8
F853
K849
β7
K812
β6
C
5'
3'
R811
G6
C9
C7
G8
H871
G8
C7
C9
G6
G10
C5
K849
C11
G4
K907
G12
C3
C13
G2
H830
K851
C14
G1
3'
5'
F853
K858
RIG-I CTD:14-bp blunt-ended dsRNA
5'
3'
G6
C9
C7
G8
G8
C7
C9
G6
K849
G10
C5
C11
G4
K907
G12
C3
C13
G2
H830
K888
C14
pppG1
3'
α
K861
β
H847
S854
F853
K858
γ
5'
RIG-I CTD:14-bp 5' ppp dsRNA

Table 3.3 Interactions between RIG-I CTD and the 14-bp non-ppp dsRNA.

Nucleotide (termini) Atoms	RIG-I CTD Atoms	Distances (Å)	Type of Interactions
G1 (5')			
O2'	His830 ND1	2.7	HB
N3	Phe853 CE2	3.2	HP
O6	Lys858 NZ	2.7	HB
O4'	Ile875 CD1	3.4	VDW
C5'	Val886 O	3.0	VDW
O5'	Gly874 CA	3.5	VDW
O2'	Tyr831 N (S30)	2.9/2.7	sHB
O2'	Ile887 O (S30)	2.9/2.8	sHB
G2 (5')			
O1P	Trp908 CD1	3.2	VDW
O2P	Lys888 CB	3.6	VDW
O4'	His830 CE1	3.6	VDW
O1P	Glu890 N (S1)*	2.8/3.2	sHB
O1P	Ser906 O (S1)	2.8/3.3	sHB
O1P	Trp908 N (S107)	2.9/2.9	sHB
C3 (5')			
O2P	Lys907 CD	3.4	VDW
O2P	Lys909 N (S46)	2.6/2.8	sHB
O1P	Trp908 N (S107)	2.8/2.9	sHB
G4 (5')			
O2P	Lys907 NZ	3.1	ES
C5 (3')			
O3'	Leu904 O(S81)	3.4/3.4	sHB
C6 (3')			
O1P	Arg811 NH2 (S61)	2.8/2.6	sHB
O1P	Ser906 O (S61)	2.8/3.3	sHB
C7 (3')			
O1P	Arg811 NH2	3.7	ES
O2P	Arg811 NE	3.0	ES
G8 (3')			
O1P	His871 NE2	4.1	ES
C11 (3')			
O1P	Lys849 NZ	3.4	ES (?)
G12 (3')			
O2P	Lys849 NZ	3.5	ES (?)
G14 (3')			
O1P	Lys851 NZ	3.7	ES
N4	Phe853 CB	3.3	HP

HB: hydrogen bonds
VDW: van der Waals contacts
HP: hydrophobic interactions
sHB: solvent mediated hydrogen bonds. Solvent distances to protein atoms and to dsRNA atoms are shown. * Solvent molecules mediating the hydrogen bonds.
ES: electrostatic interactions

The structure of RIG-I CTD shows shape and charge complementarity with the first-turn of the blunt-ended dsRNA (Figure 3.3). The average Sc of 0.67 between the two interfaces shows a better match compared to the interface between RIG-I CTD and 5'-ppp dsRNA (Sc = 0.62, Chapter II), where an Sc value of 1.0 indicates a perfect fit between two interacting surfaces (96).

The 5' four nucleotides of the dsRNA make major contributions to RNA binding, contributing to about 770 $Å^2$ of buried surface area (Figure 3.4A). However, unlike similar dsRNA with 5'-ppp, the complementary strand also makes significant contribution to the binding (buried surface area of about 540 $Å^2$) (Figure 3.4B and C). The total buried surface area at the non-ppp dsRNA:RIG-I CTD interface (about 1310 $Å^2$) is slightly larger than that at the 5'-ppp dsRNA: RIG-I CTD interface (about 1200 $Å^2$, Chapter II).

Interactions between RIG-I CTD and the non-ppp dsRNA

The interaction between the blunt-ended non-ppp dsRNA and RIG-I CTD contains a mixture of electrostatic, direct and solvent mediated hydrogen bonding, hydrophobic and van der Waals interactions (Table 3.3).

The first four nucleotides at the 5' end of the dsRNA interact with RIG-I CTD in a similar manner as dsRNA containing 5'-ppp (Figure 3.4A). The 2'-hydroxyl of nucleotide G1 forms a hydrogen bond with the side chain of His830. This is the only direct contact between the ribose and the protein. This hydroxyl group also forms two solvent-mediated hydrogen bonds with backbone amine and carbonyl groups of Tyr831

and Ile887. This ribose also has van der Waals contacts with two hydrophobic residues Val886 and Ile875. The second nucleotide (G2) interacts with RIG-I CTD through three solvent-mediated hydrogen bonds with residues Ser906, Trp908, and Glu890. G2 also has van der Waals contacts with Lys888 and His830. The third nucleotide C3 interacts with RIG-I CTD mainly through solvent mediated hydrogen bonds between its phosphodiester and the backbone amines of Trp908 and Lys909. The side chain amine of Lys907 interacts with the phosphodiester of G4 through electrostatic interactions (Figure 3.4A and C). Although similar interactions were also observed in the complex between 5'-ppp dsRNA and RIG-I CTD, the 5'-ppp makes additional electrostatic interactions with RIG-I CTD (Figure 3.4C).

Several nucleotides from the complementary strand also make significant contributions to blunt-ended dsRNA binding (Figure 3.4B and C). Key contributions likely come from positively charged Arg811 and His871 that interact with the phosphodiester of C7 and G8 through electrostatic interactions. Notably, these interactions are not observed in the structures of 5'-ppp dsRNA bound to RIG-I CTD (Figure 3.4C). The side chain of Lys849 is within 4 Å of the phosphodiester of C11 and G12, and might interact with the RNA through additional electrostatic interactions. Moreover, the side chain amine of Lys851 is 3.7 Å away from the phosphate of C14 and interacts with the RNA through electrostatic interactions. Apart from these electrostatic interactions, three solvent mediated hydrogen bonds are observed between nucleotides C5 and G6 and residues Leu904, Arg811 and Ser906.

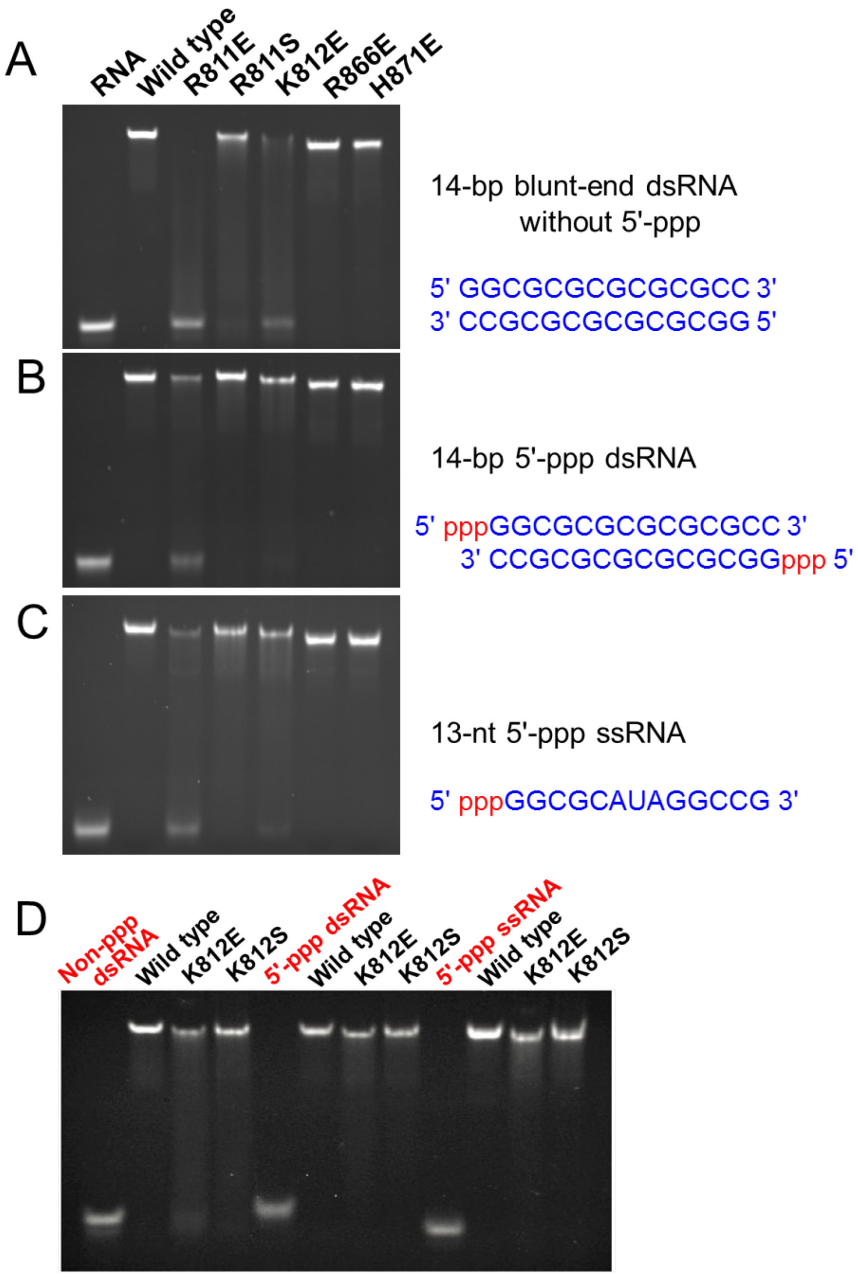


Figure 3.5 RNA binding studies of RIG-I CTD mutants by EMSA.

(A) Binding studies of wild-type and mutants of RIG-I CTD with the 14-bp blunt-ended non-ppp dsRNA.

(B) Binding studies of wild-type and mutants of RIG-I CTD with a 14-bp 5'-ppp dsRNA.

(C) Binding studies of wild-type and mutants of RIG-I CTD with a 13-nt 5'-ppp ssRNA.

(D) Binding studies of RIG-I CTD mutants K812E and K812S by EMSA.

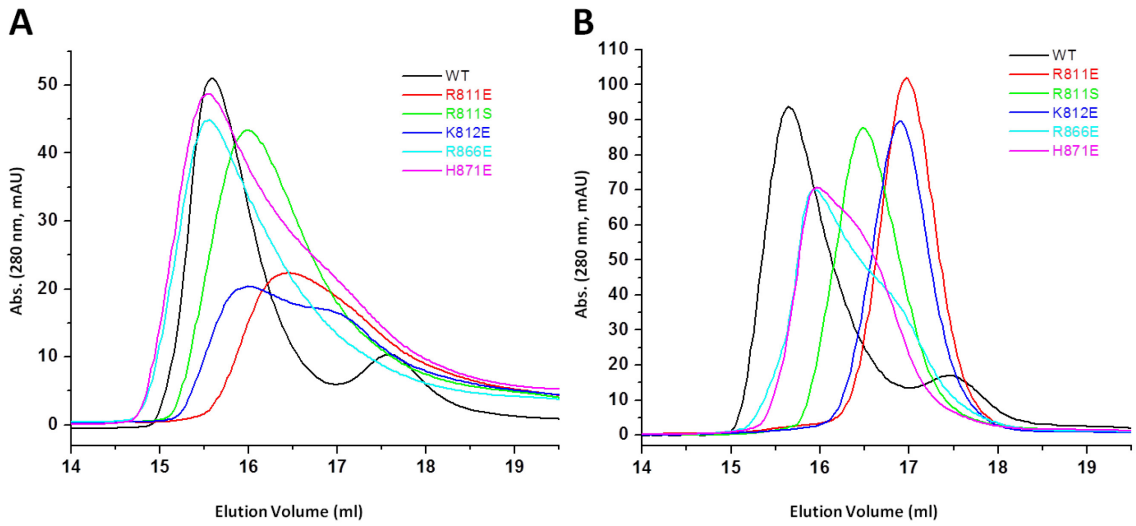


Figure 3.6 RNA binding studies of RIG-I CTD mutants by gel filtration chromatography. (A) Binding studies of wild-type and mutants of RIG-I CTD with the 14-bp 5'-ppp dsRNA.

(B) Binding studies of wild-type and mutants of RIG-I CTD with the 14-bp blunt-ended non-ppp dsRNA.

The exposed GC base pair at the termini of the dsRNA interacts with RIG-I CTD primarily through hydrophobic interactions by stacking its guanine ring on the phenyl side chain of Phe853 (Figure 3.4A). The guanine base also forms a hydrogen bond with the side chain of Lys858. A purine base at the 5'-terminus of the dsRNA should stack more effectively on the phenyl ring of Phe853 than a pyrimidine base.

Mutagenesis studies of non-ppp dsRNA binding surface of RIG-I CTD

The structure of RIG-I CTD bound to the blunt-ended non-ppp dsRNA indicates that residues Arg811, Lys812 and His871 may play critical roles in the specific binding of non-ppp dsRNA. To test this prediction, we purified RIG-I CTD mutants R811E, R811S, K812E, K812S, H871E, and R866E as a control, and conducted RNA binding studies by EMSA (Figure 3.5) and gel filtration chromatography (Figure 3.6).

Substitution of Arg811 with glutamate almost abolished blunt-ended non-ppp dsRNA binding and significantly reduced binding to both 5'-ppp dsRNA and ssRNA (Figure 3.5). Replacement of Arg811 by a hydrophilic serine reduced the binding of blunt-ended dsRNA, but did not affect binding to 5'-ppp dsRNA or ssRNA (Figure 3.5). Substitution of the nearby Lys812 by glutamate dramatically reduced non-ppp dsRNA binding, but only slightly reduced binding to 5'-ppp dsRNA or ssRNA (Figure 3.5). Replacement of Lys812 by a neutral serine residue did not affect RNA binding (Figure 3.5D). Surprisingly, the replacement of His871 by glutamate did not affect the binding to all three forms of RNA (Figure 3.5), suggesting that the His871 makes only minimal contribution to RNA binding. As expected, the replacement of Arg866, a residue not at

the RNA binding surface, by glutamate did not affect RNA binding (Figure 3.5). The slightly slower migration of the R866E and H871E complex is likely due to the changes of the surface electrostatics of the mutant proteins.

In addition, studies in Chapter II showed that mutation K907E disrupted non-ppp dsRNA as well as 5'-ppp dsRNA binding (Figure 2.12). Together, these results suggest that the electrostatic interactions between Arg811 and Lys907 and the dsRNA are essential for non-ppp dsRNA binding. Furthermore, studies in Chapter II showed that substitution of Phe853 by serine also reduced dsRNA binding (Figure 2.12), suggesting the hydrophobic interaction between Phe853 and the exposed base pair is also needed for effective RNA binding.

Blunt-ended dsRNA with and without 5'-ppp bind RIG-I CTD in different orientations

The crystal structures of dsRNA with and without 5'-ppp bound to RIG-I CTD (Chapter II and this chapter) revealed that RIG-I CTD binds 5'-ppp and non-ppp dsRNA in dramatically different ways (Figure 3.7). The non-ppp dsRNA needs to swing towards RIG-I CTD by about 15° relative to the 5'-ppp dsRNA to increase its contacts with the protein. Surprisingly, the interactions between the 5' four nucleotides of the non-ppp dsRNA and RIG-I CTD are similar to those observed in the structures of 5'-ppp dsRNA bound to RIG-I CTD. No rotation along the helical axis of the dsRNA is necessary for RIG-I CTD to interact with these two different forms of dsRNA. It is obvious that the specific interactions between the ppp and RIG-I CTD determine how 5'-ppp dsRNA

binds to RIG-I CTD, while the interactions between the phosphodiester backbones of the dsRNA and the protein determine how the non-ppp dsRNA binds.

Superposition of RIG-I CTD in the two complexes showed that the structures of RIG-I CTD are almost the same (r.m.s.d. 0.5 Å) (Figure 3.7), indicating only minor conformational adjustment of RIG-I CTD is needed to bind these two forms of RNA. These structures show that RIG-I CTD is a versatile RNA binding module that uses similar binding surfaces to recognize different forms of RNA.

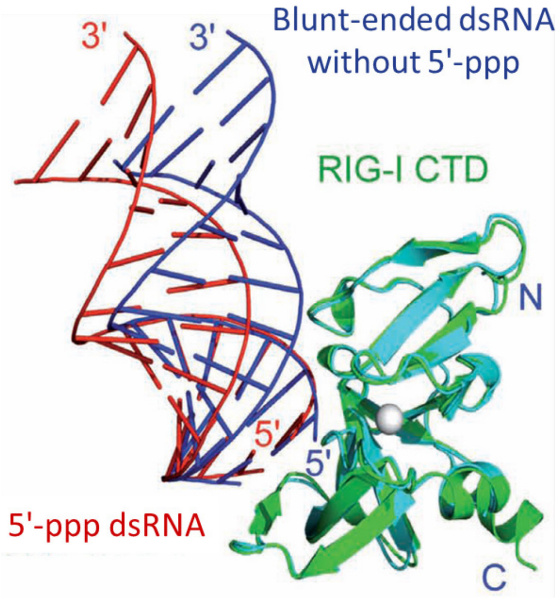


Figure 3.7 Superposition of RIG-I CTD bound to 14-bp blunt-ended dsRNA with and without 5'-ppp.
RIG-I CTD bound to the non-ppp dsRNA (blue) and the 5'-ppp dsRNA (red) are shown by the green and cyan ribbons, respectively.

Comparison of RIG-I and LGP2 CTDs bound to blunt-ended dsRNA

Superposition of the structures of RIG-I and LGP2 CTDs bound to blunt-ended non-ppp dsRNA showed that the non-ppp dsRNA bind to the two CTDs in significantly different orientations as well (Figure 3.8). The interaction between LGP2 CTD and the dsRNA is more extensive (buried surface area of about 1540 $Å^2$) and shows a slightly higher shape complementarity (Sc = 0.70) (74). To bind the LGP2 CTD, the dsRNA needs to rotate along its helical axis and swing towards the protein by about 15° relative to the position of the non-ppp dsRNA bound to RIG-I CTD (Figure 3.8). Even though the overall structures of RIG-I and LGP2 CTDs bound to dsRNA are similar (r.m.s.d. 1.35 Å), the conformations of loop5-6 that contact the exposed base pairs are quite different (Figure 3.8).

The surface electrostatics of RIG-I and LGP2 CTDs are also dramatically different (Figure 3.9). Critical residues Arg811 and His871 involved in blunt-ended non-ppp dsRNA binding by RIG-I CTD are replaced by hydrophobic Ile557 and negatively charged Glu617 in LGP2 CTD (Figure 3.9). These structural differences resulted in the dramatically different mode of non-ppp dsRNA binding by RIG-I and LGP2 CTDs.

The overall structures of RIG-I, LGP2 and MDA5 CTDs are highly conserved (r.m.s.d. 1.5 Å) (Figure 1.3 and 3.8). The largest structural differences occur at the N-terminal or C-terminal regions, which are quite flexible. The loops (loop5-6) connecting strands β5 and β6 also show high degree of flexibility. The conformation of this loop may determine how the RLRs recognize the exposed base pair at the termini of the dsRNA and likely influence the orientation of the RNA bound to the RLRs. Whether

MDA5 CTD uses a similar binding surface as RIG-I and LGP2 to bind blunt-ended dsRNA needs to be explored.

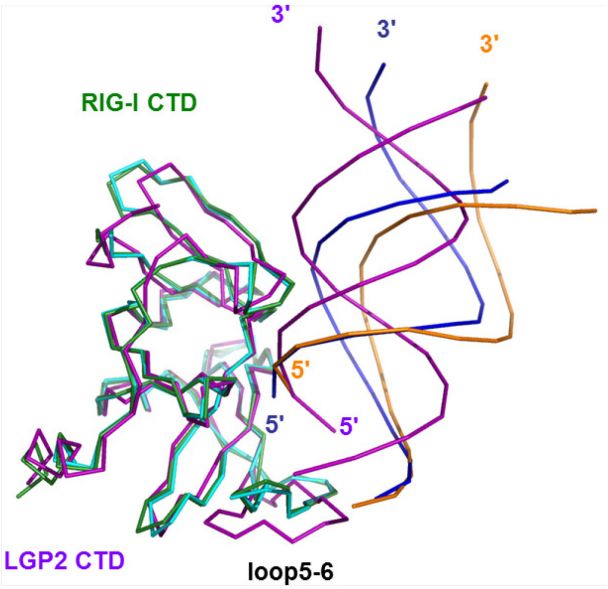


Figure 3.8 Superposition of RIG-I CTD bound to the blunt-ended dsRNA with and without 5'-ppp, and LGP2 CTD bound to blunt-ended non-ppp dsRNA.
The RIG-I CTD bound non-ppp dsRNA is shown by the blue ribbons, while the 5'-ppp dsRNA is orange. A 14-bp dsRNA (magenta ribbons) is superimposed on the 8-bp dsRNA in the LGP2 CTD:dsRNA complex structure to facilitate comparisons.

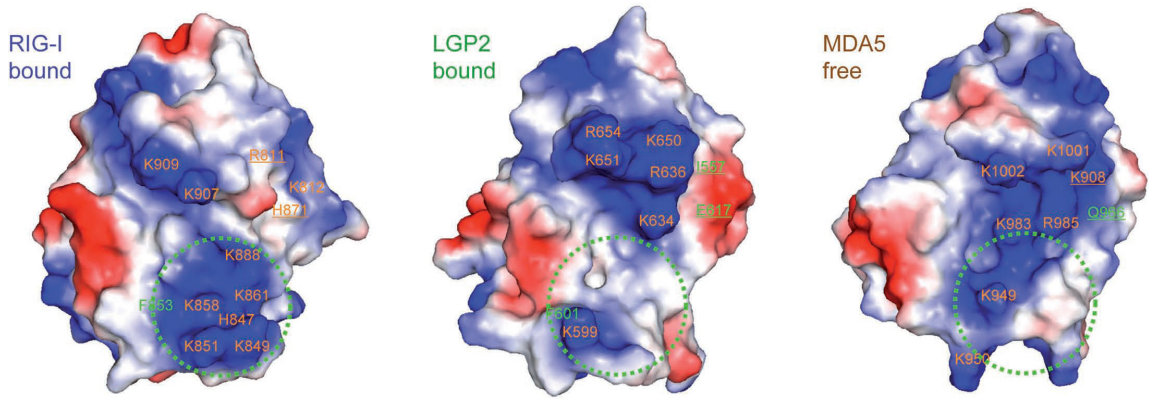


Figure 3.9 Electrostatics of the RNA-binding surfaces of the RLR CTDs. Positively charged surfaces are colored blue and negatively charged surfaces are red. The ppp-binding site of RIG-I CTD and corresponding regions of LGP2 and MDA5 are indicated by the green circles. Key residues involved in RNA-binding are labeled. Residues of RIG-I CTD interacting with the complementary strand of the dsRNA and corresponding residues in MDA5 and LGP2 CTDs are underlined.

Mutations at the RNA binding surface affect RIG-I signaling

As we discussed in Chapter II, all the three forms of RNA that bind to RIG-I CTD, including the blunt-ended dsRNA with and without 5'-ppp, are capable of activating RIG-I-dependent IFN-β reporter activities in HEK 293T cells (Figure 2.5).

To test how mutations of key residues at the dsRNA binding surface affect RIG-I signaling, we generated five mutants in the context of full-length RIG-I and conducted IFN-β luciferase reporter gene assays (Figure 3.10). Expressions of wild-type and mutants of RIG-I in the transfected cells were confirmed by western blot (Figure 3.10, inset).

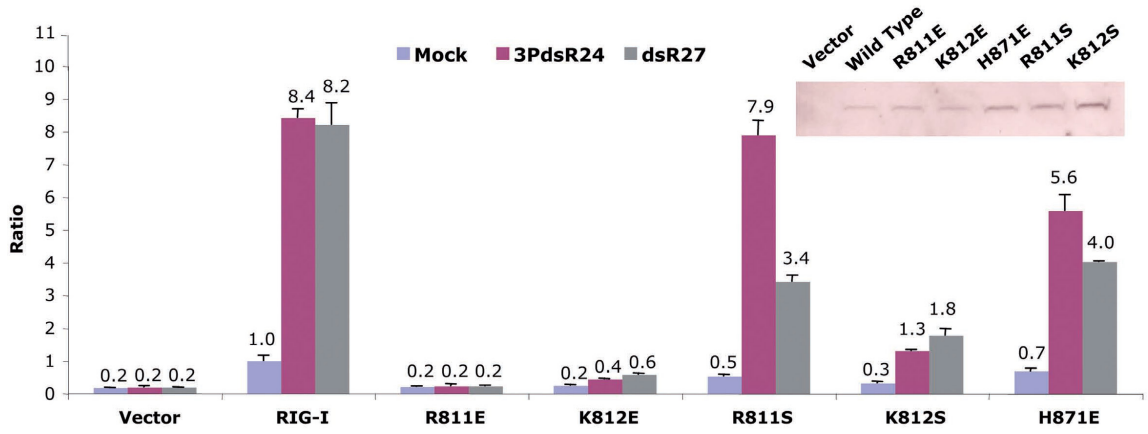


Figure 3.10 Mutations at the RNA binding surface affect RIG-I signaling.
Effects of mutations at the RNA binding surface on IFN-β luciferase production were examined in cells transfected with a 27-bp blunt-ended non-ppp dsRNA (dsR27) and a 24-bp dsRNA with 5'-ppp (3PdsR24). Expression of the wild-type and mutants of RIG-I in the transfected cells were confirmed by western blot (inset).

Substitution of Arg811with glutamate abolished RIG-I signaling for both non-ppp dsRNA and 5'-ppp dsRNA. Although the replacement of Lys812 by glutamate does not disrupt 5'-ppp dsRNA binding by RIG-I CTD, this mutation abolished signaling for both non-ppp dsRNA and 5'-ppp dsRNA. However, substitution of His871 with glutamate only reduced RIG-I signaling to 5'-ppp dsRNA by about 30%, but reduced signaling by non-ppp dsRNA by almost 50%. Consistent with results from RNA binding studies by RIG-I CTD mutants, replacement of the positively charged Arg811 with serine only slightly reduced RIG-I signaling by 5'-ppp dsRNA, but reduced the signaling

by non-ppp dsRNA by nearly 60%. In contrast, a serine substitution of the nearby Lys812 reduced signaling by non-ppp dsRNA and 5'-ppp dsRNA by about 80%.

These results demonstrated that the two positively charged residues, Arg811 and Lys812, play critical roles in the recognition of non-ppp dsRNA by RIG-I. In addition, studies in Chapter II demonstrated that substitutions of any of the three positively charged residues, Lys888, Lys907, or Lys909, by glutamate also abolished RIG-I activation by both non-ppp dsRNA and 5'-ppp dsRNA (Figure 2.14). Moreover, hydrophobic interactions between the exposed base pair and RIG-I CTD are also important for the function of RIG-I; substitution of Phe853 by serine severely decreased RIG-I signaling in response to all the three forms of RNA (Figure 2.14).

Discussion

To elucidate the structural features of viral RNA recognized by RIG-I, we expressed human RIG-I CTD, conducted RNA binding studies, and in here and Chapter II, determined the structures of RIG-I CTD bound to two different forms of RNA. These studies demonstrated that RIG-I CTD binds to 5'-ppp dsRNA or ssRNA as well as blunt-ended dsRNA without 5'-ppp. The crystal structures of RIG-I CTD bound to the 14-bp blunt-ended dsRNA with and without 5'-ppp revealed that distinct but overlapping sets of residues contact the RNA. Mutations of key residues at the RNA binding surface differentially affect RNA binding and signaling by RIG-I. These results demonstrated that RIG-I CTD is a versatile RNA binding module capable of recognizing various RNA structures. The structures of RIG-I, LGP2 and MDA5 CTDs in isolation (60,72,73) and

the structures of RIG-I and LGP2 CTDs bound to RNA (74) allowed us to compare the RNA binding properties of the RLRs.

A key question about the functions of the RLRs is what structural features of RNAs are recognized by these proteins. Biochemical, biophysical and cell-based approaches have been employed to elucidate the structures of RIG-I agonists. Several different techniques such as gel filtration chromatography, EMSA, and SPR have been used to study the RNA binding properties of RIG-I. These studies clearly showed that RIG-I CTD binds to 5'-ppp ssRNA or dsRNA, and blunt-ended dsRNA without 5'-ppp, but does not associate with ssRNA without ppp or dsDNA (74). Non-ppp dsRNA with 5'-overhangs failed to bind RIG-I CTD, while dsRNA with 3'-overhangs retained some binding. In contrast, 5'-ppp dsRNA with either 3'- or 5'-overhangs still bind to RIG-I CTD (Chapter II). Although the RNA:DNA hybrid still binds to RIG-I CTD, it forms a 1:1 instead of 1:2 complex with RIG-I CTD, suggesting that RIG-I CTD only recognizes one terminus of the hybrid with the RNA at the 5'-end. It is likely that RIG-I CTD does not bind the terminus of the hybrid with the DNA at the 5'-end. The affinity of 5'-ppp dsRNA (about 0.3 nM) for RIG-I CTD is higher than that of 5'-ppp ssRNA (about 5 nM) or non-ppp dsRNA (about 8 nM) (Chapter II). The affinities of RIG-I CTD for blunt-ended dsRNA with and without 5'-ppp determined by isothermal titration calorimetry (ITC) are significantly lower than those determined by SPR (102). This is likely due to the techniques used and the salt concentrations in the samples. These studies demonstrated that blunt-ended non-ppp dsRNA and 5'-ppp dsRNA or ssRNA are all

ligands for RIG-I. Consistent with these results, all three forms of RNA activate RIG-I signaling in cells (90,101).

The structures of RIG-I CTD bound to blunt-ended non-ppp dsRNA and 5'-ppp dsRNA provided insight into how RIG-I recognizes different forms of RNA. The two structures revealed that overlapping sets of residues are involved in the recognition of these two forms of RNA. The first four nucleotides at the 5'-end of 5'-ppp dsRNA make major contributions to 5'-ppp dsRNA binding. The ppp interacts with a cluster of positively-charged residues including Lys851, Lys858, Lys861, His847, and Lys888 through multiple electrostatic interactions. The phosphodiester backbone makes additional interactions with the protein through a combination of electrostatic, hydrogen bonding and van der Waals interactions (Chapter II). Based on the structure of RIG-I CTD bound to 5'-ppp dsRNA, removal of the triphosphate should dramatically reduce dsRNA binding. However, RIG-I CTD can bind non-ppp dsRNA with nanomolar affinity. This is likely achieved through additional electrostatic interactions between the complementary strand and Arg811 and His871 in RIG-I CTD. Consistent with this observation, replacement of Arg811 with glutamate had severe effects on non-ppp dsRNA binding, but had a more modest effect on the binding of 5'-ppp dsRNA or ssRNA binding. Moreover, substitution of Arg811 with serine only reduced the signaling of non-ppp dsRNA.

The 2'-hydroxyl of the ribose makes limited contributions to dsRNA binding. Only the 2'-hydroxyl of the 5'-terminal nucleotide is involved in direct and solvent mediated hydrogen bonds with RIG-I CTD. Modification of this hydroxyl by

methylation reduced the activity of 5'-ppp dsRNA in stimulating RIG-I signaling (102). It is likely that the steric hindrance caused by this modification reduced RNA binding. However, it is not clear why methylation of the 2'-hydroxyl of the second nucleotide also reduced RIG-I signaling (102), since this ribose makes no direct contact with RIG-I CTD.

RIG-I CTD in complex with the blunt-ended non-ppp dsRNA crystallized in a different crystal form compared to the complex with corresponding 5'-ppp dsRNA. The non-ppp dsRNA complex did not crystallize under the conditions of the 5'-ppp dsRNA complex, and *vice versa*. The crystal packing for the two complexes is significantly different. This suggested that the two forms of dsRNA bind to RIG-I CTD in different orientations in solution. Although the ppp can be flexible in solution, it has to adopt a specific conformation to bind RIG-I effectively, so it is unlikely that 5'-ppp dsRNA will bind to RIG-I CTD in a similar orientation as the non-ppp dsRNA. This orientation may reduce the specific interactions between the ppp and RIG-I CTD.

Crystal structures of RIG-I and LGP2 CTDs bound to dsRNA showed that the two proteins use similar binding surfaces to interact with RNA (Figure 3.8). It was suggested that a similar binding surface is also involved in RNA binding by MDA5 (73). Like RIG-I, LGP2 CTD also bind to the termini of the blunt-ended dsRNA (74). Some major differences in RNA binding by the three RLRs should be emphasized, however. Although LGP2 CTD binds to dsRNA with 3'-overhangs, it does not bind to dsRNA with 5'-overhangs (74). The RNA binding surfaces of the three proteins exhibit dramatically different electrostatic potentials (Figure 3.9) that should contribute to

different modes of protein:RNA interactions. The cluster of positively charged residues around Lys861 and Lys858 involved in 5'-ppp recognition is only observed at the RNA-binding surface of RIG-I (Figure 3.9), explaining why both LGP2 and MDA5 CTDs have no preference for RNA ligands with 5'-ppp. The differences in surface electrostatics also explain why RIG-I and LGP2 CTDs bind to dsRNA in different ways. The differences should contribute to the function of these three innate immune receptors in the detection of viral infections.

CHAPTER IV

INVESTIGATION OF RNA BINDING SURFACE OF MDA5*

Background

MDA5 is a second member of the RLR family proteins which is homologous to RIG-I. Both RIG-I and MDA5 encode tandem CARDs on their N-termini followed by a DExD/H box RNA helicase domain and a CTD which plays primary role in RNA recognition (49). Studies of RIG-I and MDA5 knockout mice demonstrated these two proteins detect different but overlapping sets of viruses, suggesting that they play different roles in antiviral immune responses (53,54,80). RIG-I discriminates between viral and host RNA through recognition of ssRNA with 5'-ppp group, a signature of viral RNA (19,61,62). In addition, chemically synthesized dsRNA and dsRNA derived from virus infected cells also activate RIG-I (41,52,53). Synthetic dsRNA mimics such as poly I:C or viral dsRNA of several kilobases in length activate MDA5 in cells (53,71). In contrast, short poly I:C of a few hundred base pairs in length only activates RIG-I (71).

---

*Part of this Chapter is reproduced with permission from "Structural basis of double-stranded RNA recognition by the RIG-I like receptor MDA5" by Li X., Lu C., Stewart M., Xu H., Strong R.K., Igumenova T., and Li P. 2009. *Archives of Biochemistry and Biophysics*. 488, 23-33. Copyright (2009) by Elsevier.

It is likely that MDA5 is responsible for the detection of long dsRNA while RIG-I is activated by short dsRNA in addition to 5'-ppp ssRNA. The molecular mechanism of how viral RNA activates the RLRs is still largely unknown.

The crystal structures of RIG-I CTD in isolation (72) and in complex with dsRNA with and without 5'-ppp (Chapter II and III) have been determined, providing insights into how RIG-I senses viral RNA. In addition, the crystal structures of LGP2 CTD in isolation (60) and in complex with an 8-bp non-ppp dsRNA (74) were also solved. Strikingly, the LGP2 CTD/dsRNA complex structure revealed that LGP2 specifically recognizes the blunt ends of dsRNA. LGP2 CTD exhibits a high degree of charge and shape complementarity to the first turn of dsRNA and interacts with the backbone of the dsRNA through extensive electrostatic interactions and hydrogen bonding. The exposed bases at the terminus of the dsRNA interact with LGP2 primarily through hydrophobic interactions. Similar to LGP2, the RIG-I CTD also recognizes blunt-ended dsRNA with and without 5'-ppp in addition to 5'-ppp ssRNA but does not bind to dsRNA with either 5'- or 3'-overhangs efficiently. Moreover, NF-κB and IFN-β reporter assays showed that RIG-I was activated by short dsRNA with blunt ends but not by dsRNA with 3'- or 5'-overhangs. Structures of RIG-I CTD complexes showed that RIG-I interacts with the termini of dsRNA with and without 5'-ppp in different orientations. However, it is not clear how MDA5 binds to RNA.

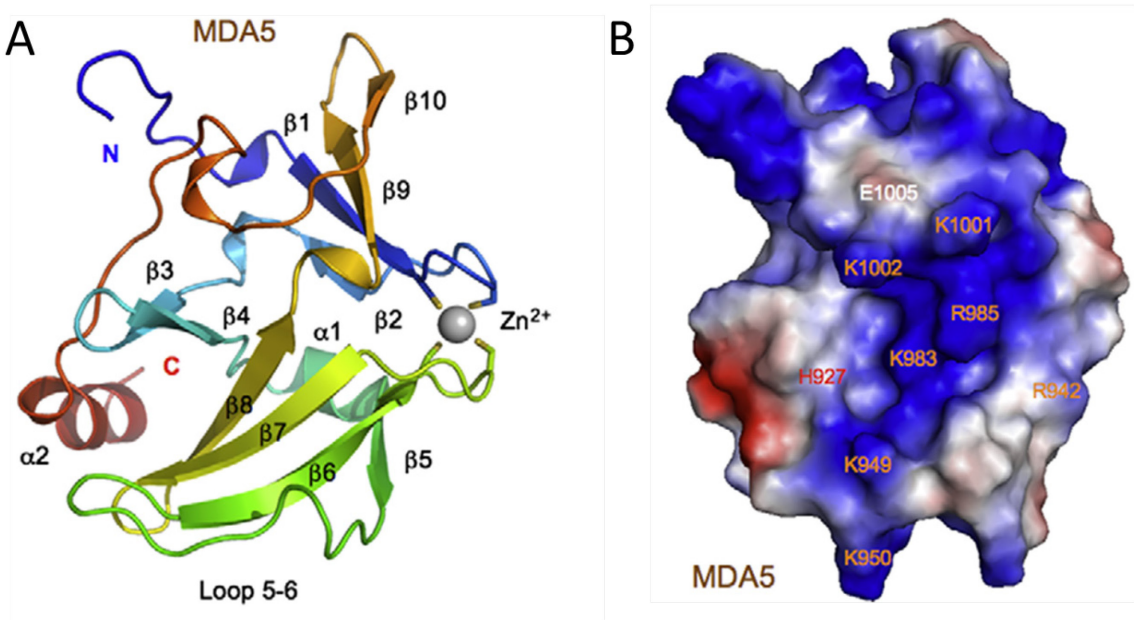


Figure 4.1 Crystal structure of human MDA5 CTD (PDB code 3GA3).
(A) Ribbon representation of the structure of MDA5 CTD. The protein molecule is colored rainbow from blue at the N-terminus to red at the C-terminus. The zinc ion is shown by the gray sphere and side chains of four cysteine residues coordinating with the zinc ion are shown by stick models.
(B) Surface electrostatic potential of MDA5 CTD. The orientation of MDA5 CTD is the same as in panel A with predicted binding surfaces facing the reader. Positively charged regions are colored blue, while negatively charged regions are red.

To elucidate the types of RNA recognized by MDA5, our lab have performed binding studies by gel filtration chromatography which showed that MDA5 CTD binds to blunt-ended dsRNA with and without 5'-ppp, and a short hairpin RNA containing only one blunt end, but not to 5'-ppp ssRNA, or dsRNA with either 5'- or 3'-overhangs (103). The stoichiometry of the dsRNA binding is estimated to be 1:1. Together with the equilibrium binding studies by SPR showing comparable affinities (K_d) of 5'-ppp dsRNA (5.7 ± 0.3 μM) and non-ppp dsRNA (2.8 ± 0.3 μM), these indicate that the 5'-ppp is not absolutely needed for dsRNA binding and MDA5 does not have a preference for the ppp group. The SPR sensorgrams showed characteristics of fast kinetics for the association and dissociation reactions, indicating that electrostatic interactions may have played a key role in the interactions between MDA5 CTD and the dsRNA (103).

To understand the structural basis of dsRNA recognition by MDA5, our lab has determined a 1.45-Å resolution structure of human MDA5 CTD (residues 892-1017 with a C-terminal 6×His tag) (Figure 4.1A). The structure of MDA5 CTD revealed a highly conserved fold similar to the RIG-I and LGP2 CTDs (52,60,72). MDA5 CTD contains a four-stranded (β1, β2, β9 and β10) antiparallel β-sheet near its N-terminus and another four-stranded (β5, β6, β7 and β8) antiparallel β-sheet in the middle. The two β-sheets are connected by a β-hairpin formed by strands β3 and β4 and a short α-helix α1. Four conserved cysteine residues, Cys907, Cys910, Cys962, and Cys964, in the two loops connecting strands β1–β2 and β6–β7 make additional connection between the two β-sheets by coordinating a zinc ion. The four cysteine residues are arranged tetrahedrally around the zinc ion and the distances between the sulfur atoms in the thiol groups and

zinc atom are about 2.35 Å. It is evident this zinc ion plays a crucial role in maintaining the overall fold of MDA5 CTD; mutations of these residues in RIG-I abrogate its response to RNA in vivo (72). Residues Tyr1015 to Glu1017 of MDA5 together with the 6×His tag form a short helix α2.

Although the amino acid sequence of MDA5 CTD is only 24.3% and 29.5% identical to the CTDs of RIG-I and LGP2, the structures of the three proteins are highly conserved, reflecting their conserved roles in dsRNA binding. The major structural differences between these proteins occur at the long loop (loop 5–6) connecting strands β5 and β6. The crystal structure of LGP2 CTD/dsRNA complex indicated that this loop is involved in the interactions between the blunt termini of dsRNA and LGP2 (74). Structural studies of RIG-I CTD by NMR spectroscopy and LGP2 CTD in isolation by crystallography indicated this loop is flexible (52,60). It is most likely that dsRNA binding induces a large conformational change in this loop to facilitate the interactions between the termini of dsRNA and RLRs.

Based on the structures of RIG-I and LGP2 dsRNA complexes, we predict that the surface defined by the β-sheet containing β5 to β8, the β-hairpin, and the three loops connecting β5–β6, β8–β9, and β9 to the C-terminal helix in MDA5 is likely involved in dsRNA binding. Examination of this potential RNA binding surface showed that it is positively charged and exhibits a high degree of shape complementarity to the structure dsRNA (Fig. 4.1B), although it is different from that of RIG-I and LGP2 CTDs, reflecting a different role in viral RNA sensing.

Because we failed to cocrystallize MDA5 CTD together with dsRNA, in order to examine the interaction between MDA5 CTD and dsRNA, we would like to explore the binding surface by NMR spectroscopy in this study.

Materials and methods

Construction of MDA5 CTD expression plasmids by molecular cloning

The cDNA of human MDA5 (IFIH1) was purchased from Open Biosystems (Thermo Scientific). The region encoding MDA5 CTD (residues 893-1017) was cloned into bacterial expression vector pET22b(+) (Novagen) between *Nde*I and *Xho*I restriction sites, in order to be expressed with an C-terminal 6×His tag. The C-terminal eight residues of MDA5 (residues 1018-1025) were excluded from the construct to prevent protein dimerization via cysteine residues 1018 and 1019. The PCR primers (IDT) used were:

Forward primer 5'- TATATATACATATGGCCAAGCATTACAAGAATAAC -3'

Reverse primer 5'- TATATACTCGAGATCCTCATCACTAAATAAAC -3'

The cloning methods were described in Chapter II. The cloned DNA sequences were confirmed by plasmid DNA sequencing (Institute of plant genomics and biotechnology, Texas A&M University).

Protein expression and purification

MDA5 CTD was expressed in *E. coli* strain BL21(DE3) by induction at OD_{600} = 0.6-0.8 with 0.1 mM IPTG overnight at 15°C. Uniform ^{13}C and ^{15}N labeling of MDA5

CTD was accomplished by growing *E. Coli* cells in minimal media M9 supplemented with 2 g/L of [^{13}C-6]-D-glucose and 1 g/L of $^{15}NH_4Cl$ (Cambridge Isotopes). The protein was purified by Ni^{2+} affinity chromatography followed by gel filtration chromatography following the same protocol for RIG-I CTD purification described in Chapter II.

NMR spectroscopy

MDA5 CTD samples for NMR experiments were prepared in 20 mM Tris-HCl pH 6.8, 150 mM NaCl, and 5 mM DTT. 8% D_2O was added to each sample for NMR lock. All NMR experiments were carried out at 25°C on Varian Inova spectrometers operating at ^{1}H Larmor frequencies of 500 and 600 MHz with help from our collaborators at Texas A&M University (Dr. Tatyana Igumenova group, Department of Biochemistry and Biophysics). Sequential assignments of the backbone ^{1}H, ^{13}Cα, ^{13}Cβ, and ^{15}N were obtained using gradient-enhanced CBCA(CO)NH and HNCACB experiments (104) with samples at 1 mM. NMR titration of MDA5 CTD with a 10-bp dsRNA was carried out at 500 MHz. The protein concentration was maintained at 100 μM, while the concentration of a 10-bp dsRNA (with palindromic sequence 5'-GGCGCGCGCC -3') was adjusted to 12, 30, 60, 120, and 240 μM in a set of NMR samples. The binding of 10-bp dsRNA to MDA5 CTD was monitored using two-dimensional ^{1}H–^{15}N HSQC spectra that correlate amide proton and nitrogen chemical shifts of the protein. NMR data were processed with nmrPipe (105) and assigned with Sparky 3 (http://www.cgl.ucsf.edu/home/sparky).

Results

^{1}H-^{15}N HSQC spectrum of MDA5 CTD

Our laboratory successfully determined the crystal structures of MDA5 CTD alone, and dsRNA complexes of RIG-I and LGP2 CTDs. Clearly, a structural comparison of the RNA complexes of RLRs would be highly informative. However, despite extensive effort, MDA5 CTD complexes resisted crystallization. As an alternative to X-ray crystallography, analysis by NMR spectroscopy in combination with molecular modeling can provide structural information, to allow comparison of the RLR CTDs interactions with viral RNA.

^{1}H-^{15}N HSQC experiment (106) is probably the most frequently recorded experiment in protein NMR spectroscopy. The resulting spectrum is two-dimensional with one axis for ^{1}H and the other for the heteronucleus ^{15}N. Each residue of the protein (except proline) has an amide proton attached to the nitrogen in the peptide bond. Each N-H pair gives an individual cross peak because of the different chemical environments. If the protein is folded, the peaks are usually well dispersed, and most of the individual peaks can be distinguished. The number of peaks in the spectrum should generally match the number of residues, with the exception of proline which lacks an amide proton hence has no cross-peaks, while residues like arginine and lysine would have additional peaks because of their side chains with nitrogen-bound protons.

Excellent chemical shift dispersion of the ligand-free MDA5 CTD ^{1}H -^{15}N HSQC spectrum (Figure 4.2) has made it possible to identify the dsRNA binding surface of MDA5 CTD. Theoretically there are 160 N-H groups in the His-tagged MDA5 CTD

protein which has 133 residues. On the HSQC spectrum, there are about 142 individual peaks. It sounds a good alternative method to study the binding surface of MDA5 CTD.

Each of the residue in MDA5 CTD was assigned with one cross peak in the HSQC spectrum (Figure 4.2) by sequential assignments of the backbone ^{1}H, $^{13}C\alpha$, $^{13}C\beta$, and ^{15}N using gradient-enhanced CBCA(CO)NH and HNCACB experiments (104).

Identification of the dsRNA binding surface of MDA5 CTD

In order to identify the residues involved in dsRNA binding surface of MDA5 CTD by NMR spectroscopy, we performed a RNA titration study by collecting HSQC spectra for a series of MDA5 CTD sample mixtures with increasing dsRNA concentrations. By overlaying each of the HSQC spectra together (Figure 4.3), we can notice the cross-peaks with chemical shift changes (e.g. decrease in intensity) which indicate residues involved in or at least perturbed by the interaction between MDA5 CTD and the dsRNA.

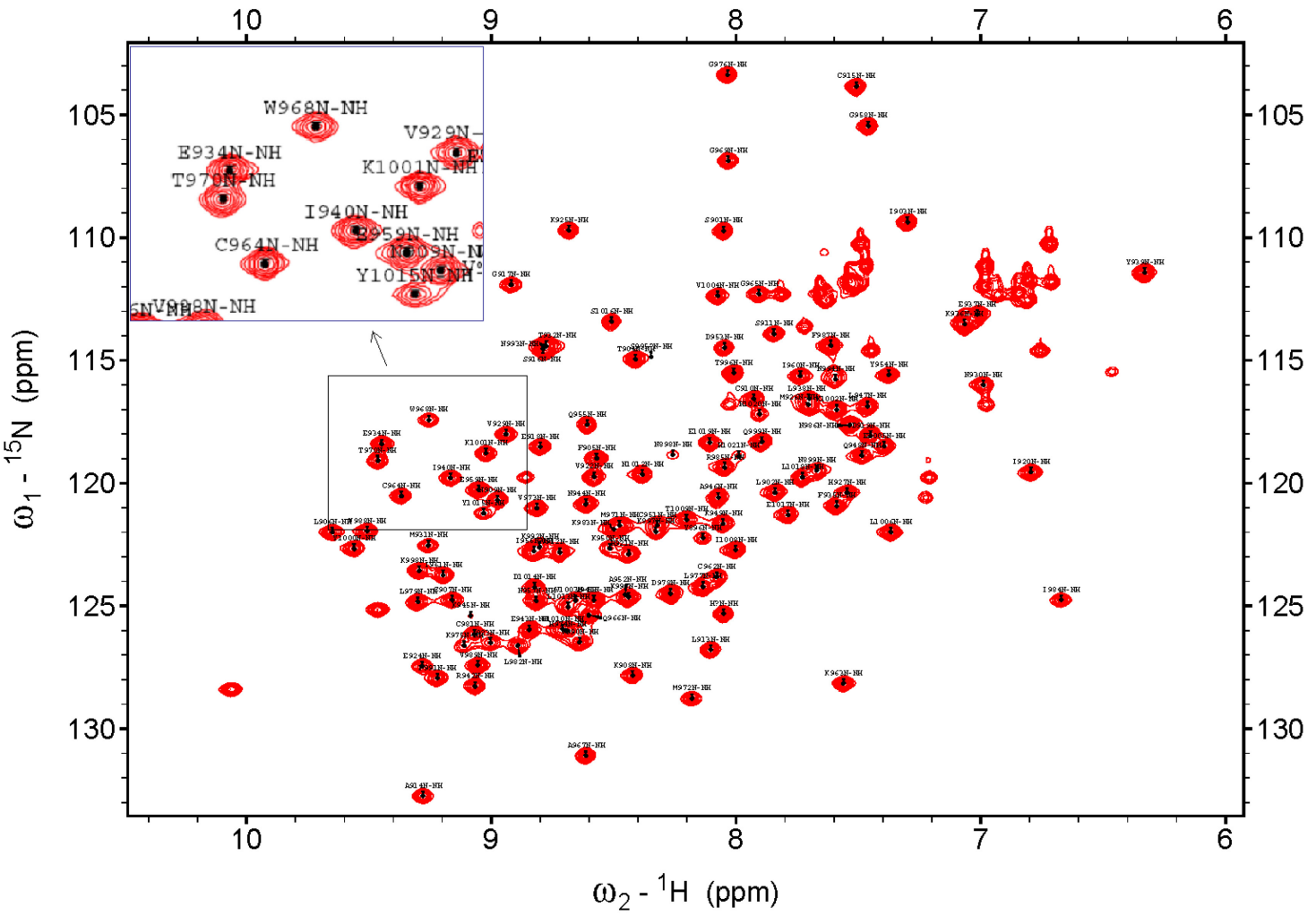


Figure 4.2 2D ^{1}H-^{15}N HSQC spectrum of MDA5 CTD with assigned cross-peak identities. Inset shows an enlarged region as an example.

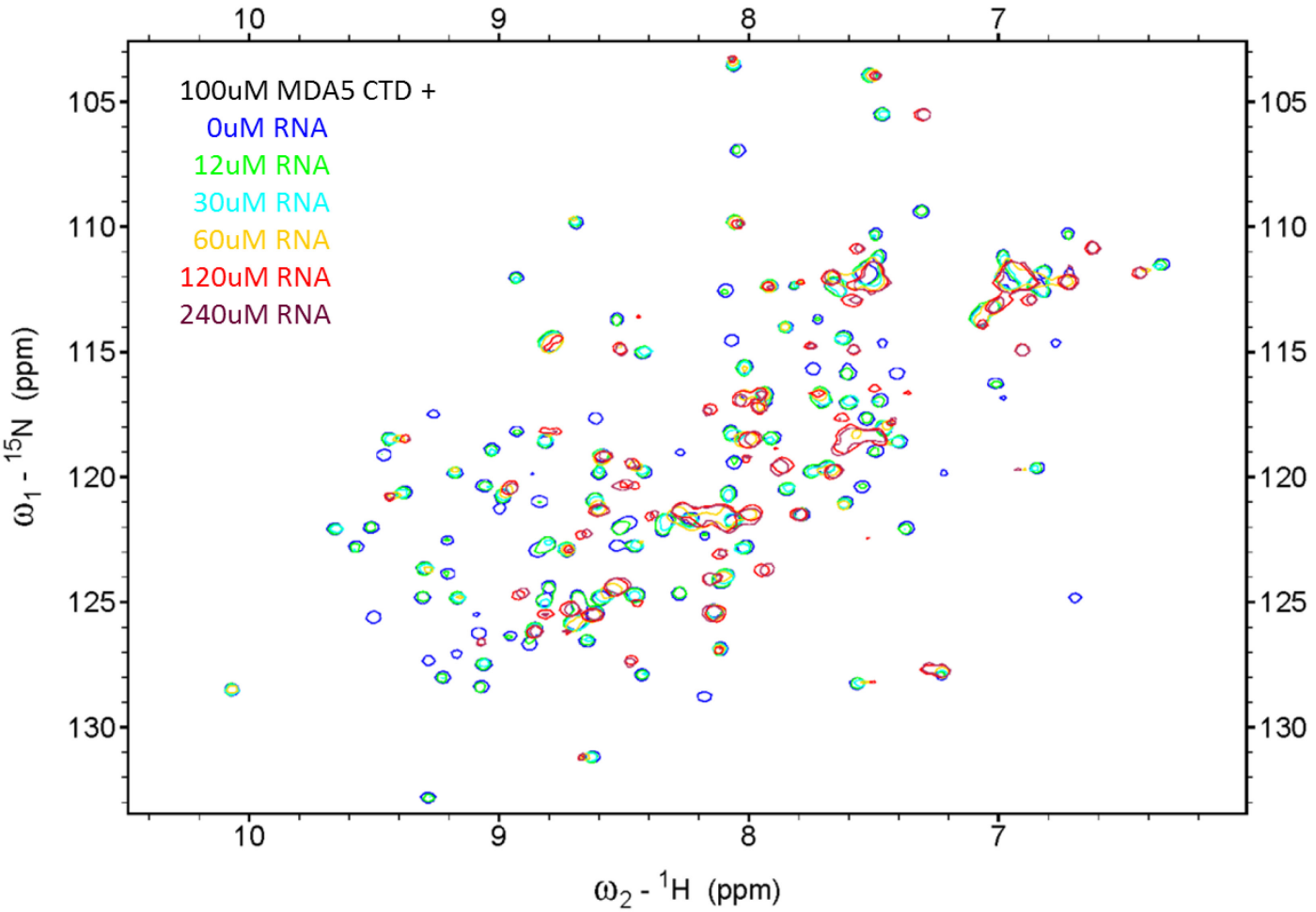


Figure 4.3 Overlay of ^{1}H-^{15}N HSQC spectra of MDA5 CTD titrated with dsRNA.

The spectra of 100 µM MDA5 CTD with 10-bp dsRNA at concentrations of 0 µM, 12 µM, 30µM, 60 µM, 120 µM, and 240 µM are shown in blue, green, cyan, yellow, red, and maroon, respectively.

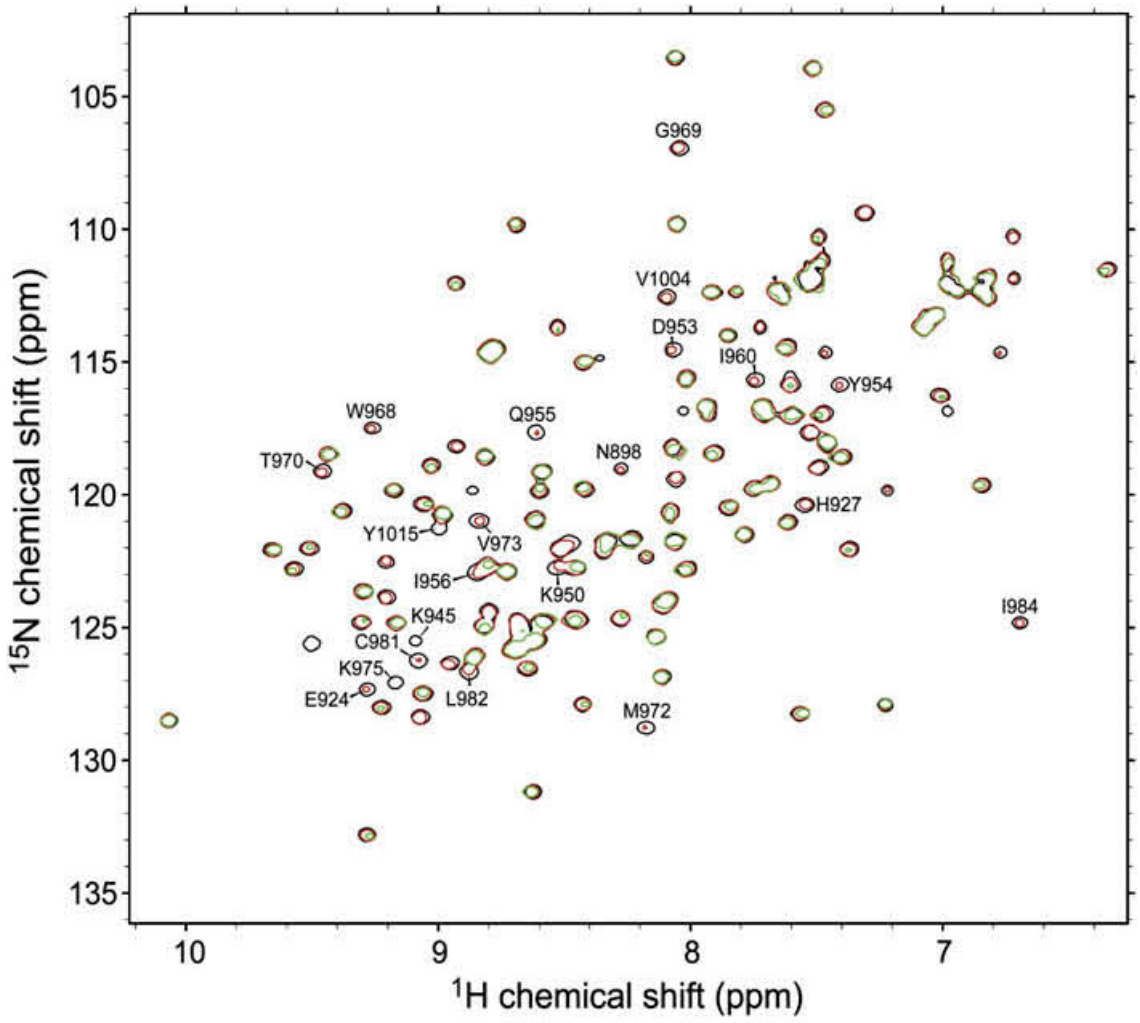


Figure 4.4 ^{1}H-^{15}N HSQC spectra of 100 μMDA5 CTD in the presence of 0 μM (black), 12 μM (red), and 30 μM (green) dsRNA.

Cross-peaks whose intensity decreased in response to adding of 12 μM dsRNA are labeled.

For the majority of residues that are involved in or perturbed by protein-dsRNA interaction, the kinetics of dsRNA binding to MDA5 CTD falls into an intermediate exchange regime on the NMR chemical shift timescale, resulting in broadening and gradual disappearance of cross-peaks with increasing ligand concentration. Even at low concentration of the dsRNA, 12 μM, the effect of ligand binding was noticeable for the 21 residues whose cross-peaks are labeled in the NMR spectrum (Fig. 4.4). At each concentration of dsRNA up to 60 μM, we identified the protein residues whose cross-peaks disappeared, or significantly changed their intensity and/or chemical shift compared to the ligand-free protein. These residues are mapped onto the crystal structure of MDA5 CTD (Figure 4.5).

Most of the residues with reduced intensity in the spectra are located on the hairpin containing strands β3 and β4, the loop connecting β5 and β6, the strands β7 and β8, the loop connecting β8 and β9, and the loop connecting strands β10 to the C-terminal helix (Figure 4.5 and 4.6). Obviously, these residues are mapped on the surface of MDA5 that corresponds to the dsRNA binding surfaces of RIG-I and LGP2 CTDs (Figure 4.6). These results indicate that all the three RLRs use a highly conserved positively charged surface to bind dsRNA. Since the overall structure of MDA5 CTD is very similar to LGP2 CTD, it is likely MDA5 binds to the blunt ends of dsRNA in a similar way as LGP2. Only a few residues on the opposite surface of MDA5 CTD showed significant changes in intensity or chemical shift upon dsRNA binding (Figure 4.5). In addition, a number of residues near the dsRNA binding surface but do not contact the dsRNA directly also showed reduced intensity in the presence of dsRNA,

indicating a global conformational adjustment of MDA5 CTD may be involved in dsRNA binding.

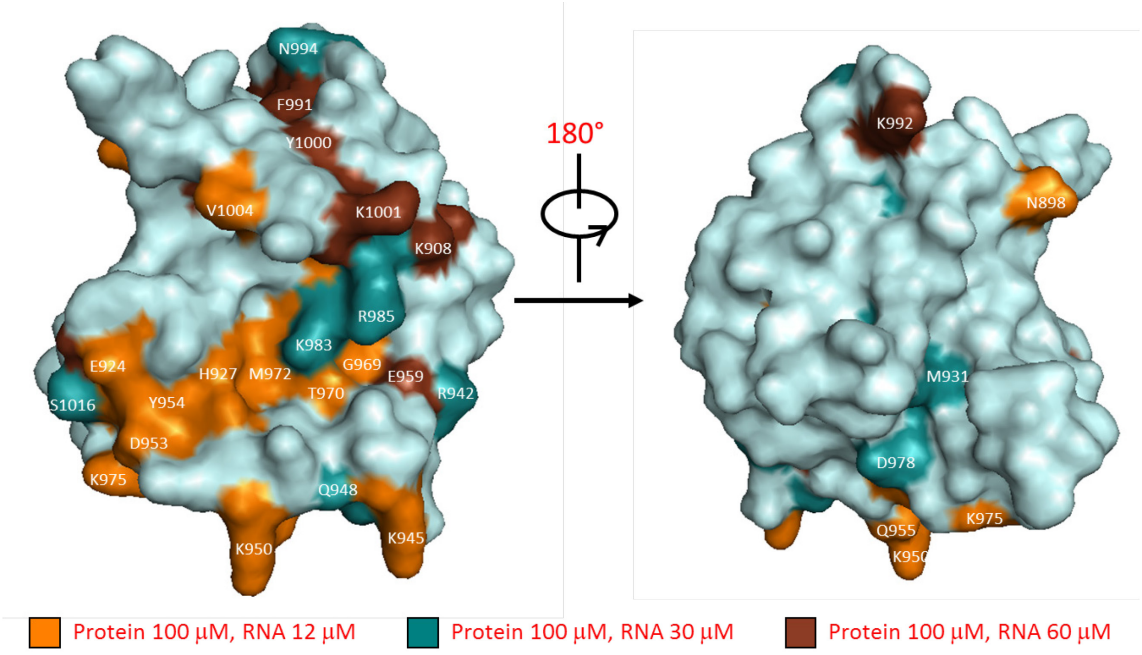


Figure 4.5 Mapping of amino acid residues involved in dsRNA binding identified by NMR spectroscopy onto the crystal structure of MDA5 CTD.
Residues involved in dsRNA binding at RNA concentrations of 12 µM, 30 µM, and 60 µM are colored orange, teal, and chocolate, respectively.

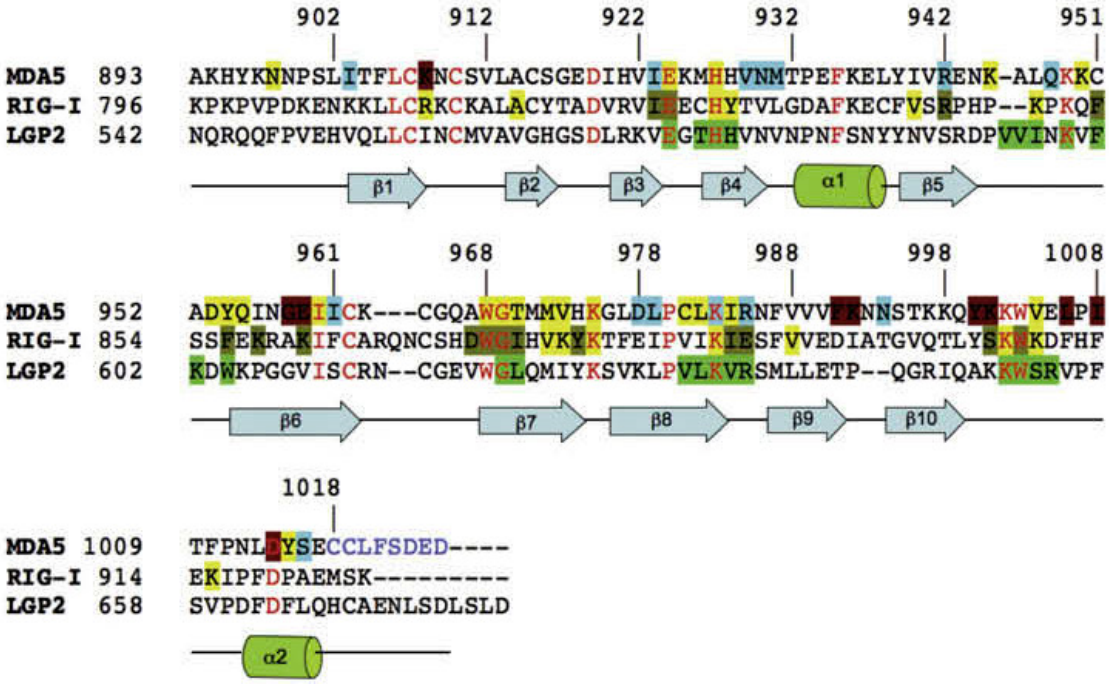


Figure 4.6 Structure-based sequence alignment of human MDA5, RIG-I, and LGP2 CTDs.

Conserved residues in the three proteins are shown in red. Residues of MDA5 not included in the protein construct are shown in light blue. Secondary structural elements of MDA5 CTD are shown under the aligned sequences. Residues of MDA5 involved in dsRNA binding at concentrations of 12 μM, 30 μM, and 60 μM are highlighted in yellow, cyan, and chocolate, respectively. NMR titration of RIG-I CTD identified overlapping sets of residues involved in dsRNA and 5'-ppp ssRNA (52). Residues of RIG-I CTD involved in dsRNA are highlighted in yellow; residues involved in 5'-ppp ssRNA recognition in addition to those residues that are involved in dsRNA binding are highlighted in olive. Residues of LGP2 CTD involved in dsRNA binding observed in the crystal structure of the LGP2 CTD/dsRNA complex (74) are highlighted in green.

HSQC spectra of MDA5 CTD-dsRNA complexes collected at 120 μM and 240 μM dsRNA concentrations are virtually indistinguishable (Figure 4.3), indicating full saturation of protein with ligand. Even at saturating concentrations of dsRNA we observed significant line-broadening of NMR cross-peaks that cannot be fully explained by an increase in molecular weight due to the protein-dsRNA complex formation. One plausible explanation is the formation of a 2:1 complex between MDA5 CTD and the dsRNA at higher concentrations, which increases the molecular weight of the complex significantly.

Discussion

RNA binding studies by gel filtration chromatography and SPR demonstrated that MDA5 CTD only binds dsRNA with blunt ends, but does not associate with dsRNA with either 5'- or 3'-overhangs. These confirmed previous findings that MDA5 is a sensor of dsRNA. Similar preferences of blunt-ended dsRNA were also observed for the LGP2 and RIG-I CTDs, indicating the blunt end of dsRNA is most likely a common structural motif recognized by the RLRs (74). However, unlike RIG-I, MDA5 and LGP2 have no preference for the 5'-ppp group. The CTD of MDA5 exhibits significantly lower affinities for dsRNA (Kd ~ 3 μM) compared to the CTD of LGP2 (Kd ~ 100 nM) (74) and RIG-I (Kd ~ 4.6 nM), suggesting that MDA5 needs higher concentrations of ligands for activation. Since MDA5 CTD has lower affinities for dsRNA, it forms a 1:1 complex with the 10-bp dsRNA at low concentration and a mixture of 1:1 and 2:1 complexes at

higher concentration, making the MDA5/dsRNA complex heterogeneous and resistant to crystallization.

To elucidate the structural basis of dsRNA recognition by MDA5, our laboratory has determined the high-resolution structure of MDA5 CTD by X-ray crystallography and in this study mapped its binding surface for dsRNA by NMR spectroscopy. The structure of MDA5 CTD revealed a highly conserved fold similar to RIG-I and LGP2 CTDs. All the three proteins show a high degree of shape and charge complementarity with the first turn of blunt-ended dsRNA. Consistent with this, NMR titration of MDA5 CTD with dsRNA indicated that a conserved positively charged surface is involved in dsRNA binding. Structural studies of RIG-I and LGP2 CTDs in complex with dsRNA by X-ray crystallography indicated that a similar binding surface is involved in dsRNA and ssRNA binding by RIG-I and dsRNA binding by LGP2.

Based on the RNA binding surface mapped by NMR spectroscopy, and the structural comparison of MDA5 CTD with the RIG-I and LGP2 CTD/dsRNA complexes, seven mutants of MDA5 CTD were generated and their binding properties were analyzed by gel filtration chromatography (103). Mutation of the positively charged residues Lys950, Lys983, and Lys1002 to negatively charged glutamate abolished dsRNA binding. Consistently, mutation of lysine residues in RIG-I and LGP2 CTDs corresponding to Lys983 and Lys1002 in MDA5 also disrupted RNA binding by RIG-I and LGP2 (52,74), demonstrating their pivotal roles in dsRNA recognition by RLRs. In addition, mutation of the conserved His927 that is involved in the formation of a network of hydrogen bonds in the RIG-I and LGP2 CTD complexes also disrupted

dsRNA binding by MDA5 CTD. However, mutation of Glu924 or Leu947 which may interact with dsRNA by hydrogen bond and hydrophobic interactions, did not abrogate dsRNA binding by MDA5 CTD, suggesting these two residues might contribute to dsRNA binding but do not play dominant roles. As a control, mutation of Arg942, a residue that is not on the predicted dsRNA binding surface of MDA5, did not affect dsRNA binding. These mutagenesis studies demonstrated that similar to RIG-I and LGP2, a highly conserved positively charged surface is involved in dsRNA binding by MDA5 CTD, and electrostatic interactions play primary roles.

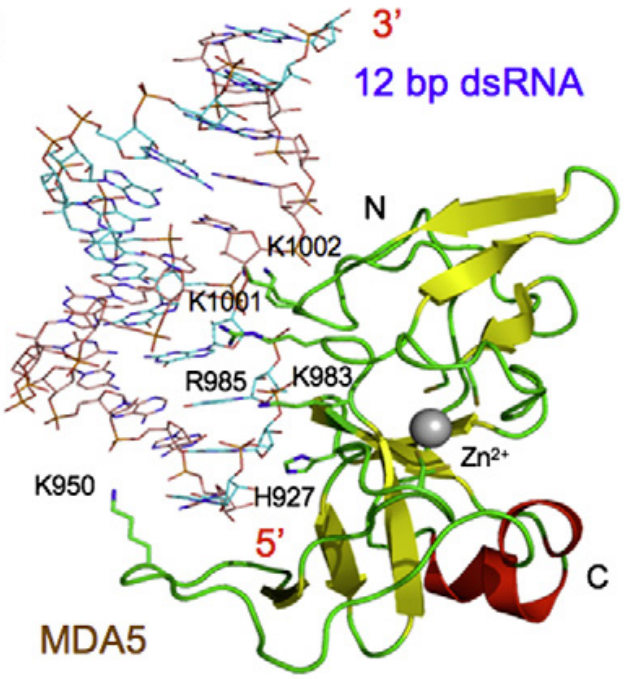


Figure 4.7 Structural model of MDA5 CTD bound to a 12-bp dsRNA.

Based on the above NMR and mutagenesis studies and other RLR CTD complex structures, a structural model of MDA5 CTD bound to a 12-bp dsRNA was generated (Figure 4.7) (103), which suggested that MDA5 may recognize the blunt end of dsRNA and bind to the first turn of it in a similar manner as its homolog LGP2.

Since several mutants of MDA5 with abolished dsRNA binding were identified, it would be interesting to test whether these mutants of MDA5 still stimulate type I IFN induction in virus infected cells. These studies will provide insight into whether the dsRNA binding is needed for MDA5 signaling. Since the coexpression of LGP2, which exhibits much higher affinities for dsRNA than MDA5, stimulated rather than suppressed the activation of MDA5 (60), the roles of dsRNA binding in MDA5 activation need to be reconsidered. It is also possible that dsRNA binding is not required for the regulation of MDA5 by LGP2. Previous studies from two independent groups already showed that mutants of LGP2 with abolished RNA binding still suppress the activation of RIG-I (74,89).

RNA binding studies clearly showed MDA5 CTD also binds to short dsRNA of 8 to 24 bp in length just like the CTDs of LGP2 and RIG-I. Since the affinity of MDA5 CTD for blunt-ended dsRNA is only about 3 μM, it is difficult to understand why full-length MDA5 prefers long dsRNA instead of short dsRNA as ligands, since the concentration of the blunt ends, the structural motif of dsRNA recognized by MDA5, would be very low for a given amount of long dsRNA. For example, a solution of a 2-kb dsRNA at 3 μM concentration corresponds to a solution of 20-bp dsRNA at 300 μM concentration with the same amount of RNA. On the other hand, the maximum length of

dsRNA that can be covered by a full-length MDA5 molecule is likely about 30 bp assuming the maximum dimension of the molecule is about 100 Å as observed in the electron microscopy structure of full-length LGP2 and RIG-I (99,100). Moreover, previous studies demonstrated that 24-bp but not 19-bp dsRNA stimulates the activation of RIG-I, indicating that the 24-bp dsRNA is long enough to form an active complex with full-length RIG-I (74). The activation of antiviral ribonuclease, RNase L, by 2',5'-linked oligoadenylate (2-5A) produces small RNA cleavage products that initiate IFN production (30). This is likely a mechanism to generate a large amount of short dsRNA to simulate the activation of the RLRs. However, it is not clear whether the products of RNase L serve as ligands for MDA5. The mechanism of how MDA5 is selectively activated by long dsRNA and poly I:C needs further investigation.

CHAPTER V

STRUCTURAL STUDIES OF RIG-I ACTIVATION

Background

The molecular mechanism of how viral RNA activates the RIG-I is still largely unknown. It was suggested that under resting condition RIG-I is in a suppressed conformation and RNA binding induces a major conformational change which requires ATP hydrolysis, exposing the CARDs for the recruitment of the mitochondria bound adaptor protein IPS-1, which relays the signal to downstream proteins (49,51). Stimulation of the RIG-I ultimately leads to the activation of transcription factors IRF3, IRF7, and NF-κB that regulate the induction of type I IFNs and proinflammatory cytokines (21,70,80).

To understand those mechanistic questions concerning RIG-I activation pathway, including the nature of the conformational changes, the role of ATP hydrolysis, and the dimerization or oligomerization state associated with viral RNA binding, it would be necessary to obtain the structural information for different states or conformations of RIG-I. Although the crystal structures of RIG-I CTD complexes with different dsRNA have been determined, the overall structure of full-length RIG-I or structures of other domains are still unknown. In this study, we would like to further investigate into the mechanism for RIG-I activation by crystallizing RIG-I HD-CTD, which contains both the helicase domain and the CTD, in complex with dsRNA.

RLRs are members of the large SF2 helicases family and each contain a DExD/H-box helicase domain. The RLRs share a conserved catalytic core consisting of two RecA-like domains similar to SF1 and SF2 helicases. Eight conserved helicase domain sequence motifs cluster into the two RecA subdomains which forms an active site for ATP binding and hydrolysis (107). The name DExD/H box refers to the signature sequence, which is "DECH" for the RLRs, in motif II which is also known as the Walker B motif that is crucial for the hydrolysis of ATP (108). Both RIG-I and MDA5 strictly require RNA ligand stimulation for the ATPase activity, while LGP2 exhibits a low level of basal ATP hydrolysis in the absence of RNA (89). However, though all three RLRs are able to hydrolyze ATP, only RIG-I has been reported to have ATP-dependent RNA unwinding activity to date (52,66). Interestingly, the helicase activity of RIG-I appears to be reversely correlated with its antiviral signaling capacity stimulated by specific RNA ligands (52). Questions remain about the roles of RNA helicase activity in RLRs activation and signaling. Therefore a structural study of RIG-I HD-CTD in complex with dsRNA would provide insights into RNA induced RIG-I activation.

Epigallocatechin-3-gallate (EGCG) is a green tea catechin that is recently identified in a small molecule screen as a RIG-I regulator that can bind to RIG-I and modulate its signaling (109). It inhibits RIG-I signaling in cell-based reporter gene assays. EGCG also inhibits the RNA-dependent ATPase activity of recombinant RIG-I in biochemical assays, however it does not compete with RIG-I interaction with RNA or

with ATP. It would be intriguing to elucidate the structural mechanism of this inhibition, which possibly involves the helicase domain, by crystallography study.

Materials and methods

Cloning and protein expression of RIG-I HD-CTD

Human RIG-I HD-CTD (residues 232-925) was cloned into the insect cell expression vector pAcGHLT-2 (BD Biosciences), which contains N-terminal GST-tag and 6×His-tag followed by a thrombin-cleavage site. Using such a construction, after thrombin cleavage, the peptide GSPGLDGSC will remain at the N-terminal end of the protein, corresponding to a molecular weight of 80196.5 Da for the isolated module. The target gene was PCR amplified and insert into the vector using *Nde*I and *Not*I restriction enzymes. The DNA sequence was verified by plasmid sequencing. The PCR primers (IDT) used were:

Forward primer 5'- TATATATACATATGCGTGTCTGATACAAAC -3'

Reverse primer 5'- TATATAGCGGCCGCTCATTTGGACATTTCTGCTGG -3'

The 6×His-GST-tagged RIG-I HD-CTD was expressed in insect cells *Sf*9 (*Spodoptera frugiperda*) using baculovirus expression system. Recombinant baculovirus was generated in *Sf*9 cells using BaculoGold linearized baculovirus DNA (BD Biosciences) and the recombinant pAcGHLT-2 plasmid according to the manufacturer's protocol. *Sf*9 cells were maintained in suspension culture in HyClone SFX-serum free insect cell culture medium (Thermo Scientific) at 27°C. For large-scale protein

expression, 2 L of *Sf*9 cell culture at a density of 2.5×10^6 ml^{-1} were infected with 120 ml recombinant virus and incubated at 120 rpm at 27°C for 58 h.

To be expressed in bacterial system, RIG-I HD-CTD was also cloned into a pET28a(+) derived vector with an N-terminal SUMO tag. The target gene was PCR amplified and insert into the vector using *Sal*I and *Not*I restriction enzymes, to be expressed with an N-terminal 6×His-SUMO fusion. The linker region between SUMO and HD-CTD were then removed by mutagenesis PCR. Using such a construction, the expressed HD-CTD protein will have no extra residues before its N-terminal after treated with SUMO protease. The PCR primers (IDT) used in this study were:

For cloning PCR

Forward primer 5'- TATATAGTCGACAAGTGTCTGATACAAACTTGTAC -3'

Reverse primer 5'- TATATAGCGGCCGCTCATTTGGACATTTCTGCTGG -3'

For mutagenesis PCR

Forward primer 5'- GAACAGATTGGTGGATCCGTGTCTGATACAAACTTG -3'

Reverse primer 5'- CAAGTTTGTATCAGACACGGATCCACCAATCTGTTC-3'

The protein was expressed in *E. coli* BL21(DE3) cells. The culture was grown at 37°C in LB medium with 50 μg/ml kanamycin to an OD_{600} of 0.6-0.8. $ZnSO_4$ (at 0.2 mM) was added to the medium because RIG-I HD-CTD is a Zinc binding protein. The culture was then induced with 0.5 mM IPTG at 15°C and grown overnight.

Purification and crystallization of RIG-I HD-CTD in complex with a 24-bp dsRNA

*Sf*9 insect cells expressing human RIG-I HD-CTD were harvested 58 h post-infection by centrifuging at 2000 rpm for 5 min. The cell pellets were resuspended in 150 ml Lysis Buffer [200 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% (v/v) NP40, and 1mM phenylmethylsulfonyl fluoride (PMSF)] and lysed at 4°C for 2 h on a rolling shaker. Cellular debris was sedimented by centrifugation of the lysate at 16000 rpm for 30 min. The supernatant was incubated at 4°C for 2 h with 6 ml Ni-NTA superflow (Qiagen) on a rolling shaker. The slurry was loaded by gravity into an empty column and washed with about 200 ml of the Washing Buffer, prior to elution with about 25 ml Elution Buffer. After eluted from the Ni-NTA beads, 5 mM DTT was added to the protein immediately.

The 24-bp non-ppp dsRNA (IDT) used to form complex with RIG-I HD-CTD was generated by annealing of two complementary ssRNAs (5'- GCGCACGUGCGC GCGCGUGCACGC and 5'- GCGUGCACGCGCGCGCACGUGCGC). The purified RIG-I HD-CTD protein was mixed with the 24-bp dsRNA at a molar ratio of about 3:1. The mixture was stored at 4°C for 2 days to form stable 2:1 (protein:RNA) complex before thrombin cleavage to remove the 6×His-GST-tag. The complex was then purified by gel filtration chromatography for twice on a Superdex 200 (1.6 × 60) column (GE Healthcare) in the Tris-NaCl buffer. The purified complex sample was concentrated to about 5 mg/ml using 30kDa cutoff Amicon Ultra centrifugal filter units (Millipore).

All crystallization experiments employed the hanging-drop vapour-diffusion method and were conducted at 4°C, using 24-well plates (Hampton Research). Initial

screening was performed against 336 crystallization conditions using Index, Crystal Screen (I & II), Natrix, and PEG/Ion (I & II) kits (Hampton Research). Drops were prepared by mixing 2 µl reservoir solution with 2 µl protein solution (5 mg/ml HD-CTD/RNA complex in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM DTT, and 1 mM ATP) and were equilibrated by vapour diffusion against 500 µl reservoir solution. Several initial hits were obtained with small needle-like crystals within a week. After extensive optimization trials, plate-like rod crystals with maximum dimensions of 300 × 100 × 20 μm^3 were obtained in 0.2 M $MgCl_2$ and 12% (w/v) PEG 3350 within two weeks.

Crystals were transferred into a cryoprotectant solution composed of reservoir solution with 30% (w/v) sucrose and then flash-frozen in liquid N_2. Diffraction data were collected at Stanford Synchrochon Radiation Lightsource (SLAC National Accelerator Laboratory, Menlo Park, CA, USA). Data were indexed, integrated and scaled using HKL-2000 (92).

Purification and crystallization of RIG-I HD-CTD in complex with a 14-bp dsRNA

E. coli BL21(DE3) cells expressing human RIG-I HD-CTD were harvested by centrifugation at 4°C, 6000 rpm for 10 min. Cell pellets were lysed and purified by nickel affinity chromatography on a His-Select nickel affinity resin (Sigma-Aldrich) column following the same protocol for RIG-I CTD purification described in Chapter II.

The 14-bp 5'-ppp dsRNA used to form complex with RIG-I HD-CTD was generated by annealing of a T7 RNA polymerase *in vitro* transcribed ssRNA with

palindromic sequence 5'- GGCGCGCGCGCGCC). The purified RIG-I HD-CTD protein was mixed with the 14-bp dsRNA at a molar ratio of about 1:1.5 to form a 1:1 (protein:RNA) complex. The 6×His-SUMO-tag was removed by SUMO protease treatment. The complex was then purified by gel filtration chromatography on a Superdex200 (1.6 × 60) column (GE Healthcare) eluted with the Tris-NaCl buffer. The purified complex sample was concentrated to about 10-15 mg/ml using 10-kDa cutoff Amicon Ultra centrifugal filter units (Millipore).

The HD-CTD/14-bp dsRNA complex (in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM DTT, 5 mM $MgCl_2$) in the presence of 1 mM ATP or inhibitor EGCG (Sigma Aldrich) was crystallized by hanging-drop vapour-diffusion method at 16°C in 17-20% (w/v) PEG 3350, 0.2 M NaSCN, and 6% trifluoroethanol. Crystals appear in 2-3 days and grow into large hexagonal rods in 1-2 weeks.

Results

Purification and crystallization of RIG-I HD-CTD in complex with a 24-bp dsRNA

We have successfully cloned the human RIG-I HD-CTD gene and generated recombinant baculovirus for the overexpression of the protein in eukaryotic *Sf*9 cells for proper protein folding. The protein/RNA complex sample for crystallization was purified to homogeneity according to the purification protocols outlined in the materials and methods section. RIG-I HD-CTD with N-terminal 6×His-GST fusion was first purified using Ni-affinity and immediately mixed with the 24-bp dsRNA to form stable complex before the GST-tag was removed, otherwise the protein tends to precipitate during buffer

exchange and thrombin cleavage. The complex was then purified by gel filtration chromatography. To completely separate the 2:1 complex from extra protein and uncut GST-tagged protein, the complex was purified two times using a Superdex 200 gel filtration column and only the best fractions were collected for crystallization experiments. Figure 5.1 represents a typical gel filtration purification profile. The purity of the protein was verified on a 10% SDS-PAGE gel showing a single band at 80 kDa (Figure 5.2).

This complex sample was used to set up crystallization screens using over 300 different conditions. Small crystals in the forms of fine needles or needle clusters appear after about one week in several conditions. Several rounds of optimization were carried out with different precipitant concentrations and pH, however the crystals were not improved much.

Next we tried to add 2 mM of ATP to our complex sample mixture for crystallization. Better crystals appear in the shape of thin plates instead of fine needles, in optimization conditions containing 0.2 M $MgCl_2$, and 10-13% PEG 3350. This indicated that the ATP hydrolysis by the helicase domain may be required for a conformational change of the RIG-I HD-CTD which stabilizes the complex and make it favorable for crystallization. Based on this, we also tried to use ADP, the nonhydrolyzable ATP analog AMP-PNP, or the transition state analog ADP-AlF_3. However, none of them resulted in better crystals.

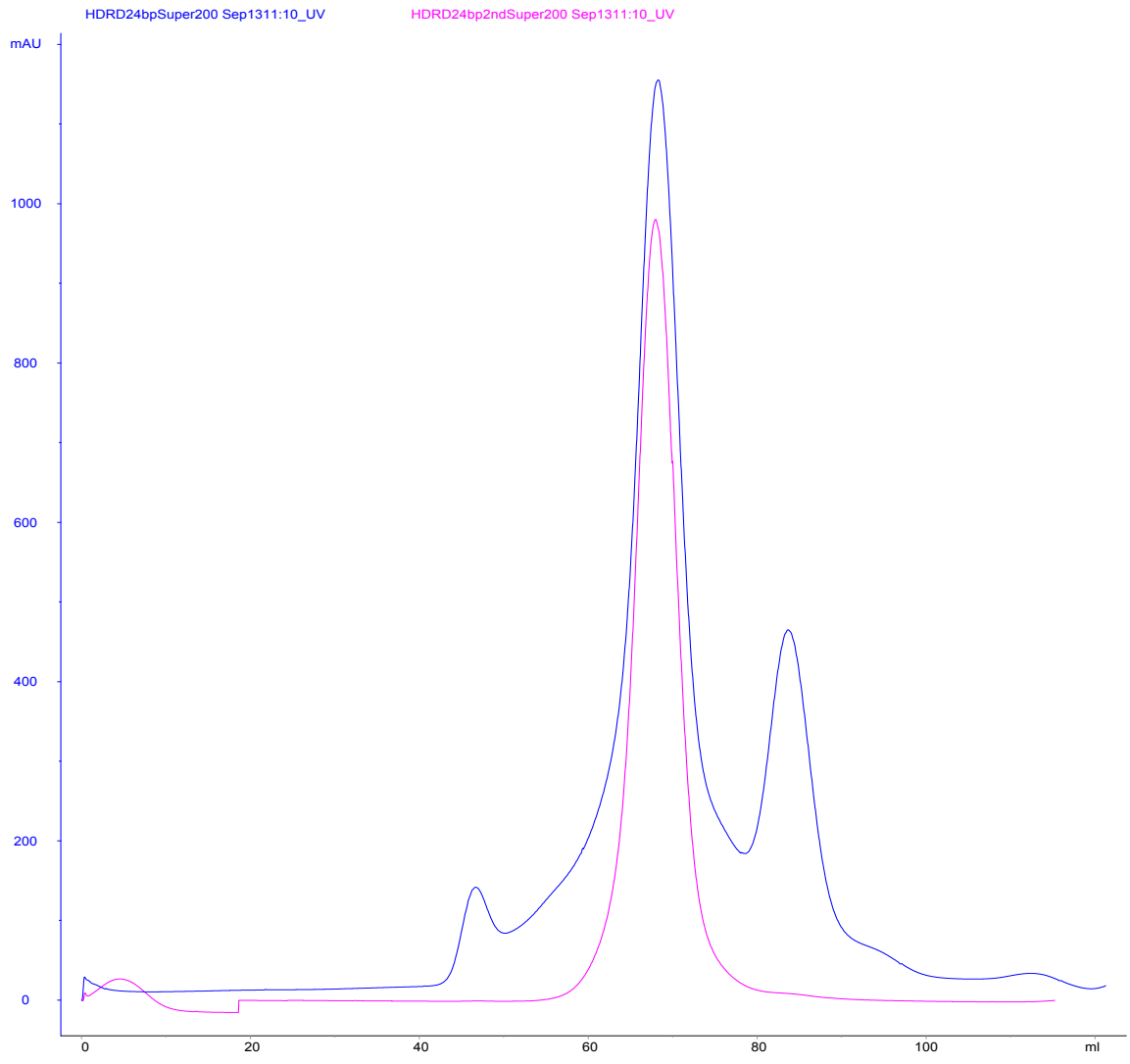


Figure 5.1 Gel filtration chromatograms of RIG-I HD-CTD complex purification. The purification profile of RIG-I HD-CTD in complex with the 24-bp dsRNA on a Superdex 200 column shows the absorbance peak at 280 nm. The first round of purification is shown in blue chromatogram and the second shown in red.

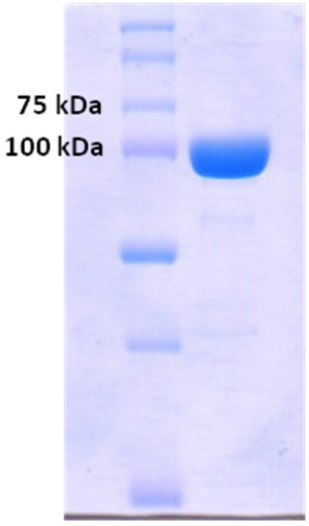


Figure 5.2 SDS-PAGE analysis of purified RIG-I HD-CTD complex.
RIG-I HD-CTD together with Precision Plus Protein Dual Color Standards (25-250 kDa, Bio-Rad) were shown on a 10% SDS-PAGE gel stained with Coomassie Blue.

After further optimization trials by changing the concentrations of protein complex, ATP, salt $MgCl_2$, and precipitant PEG 3350, plate-like crystals with maximum dimensions of 300 × 100 × 20 μm^3 were obtained using 5 mg/ml in the presence of 1 mM ATP, in 0.2 M $MgCl_2$ and 12% (w/v) PEG 3350. The crystals were harvested and subjected to X-ray diffraction studies using synchrochon beamline. The crystals diffracted to 3.5-Å resolution. However, the diffraction data showed that the crystals were not single and cannot be used for structure determination. Preliminary analysis of the diffraction data indicated that the crystals belong to the space group P222, with unit-

cell parameters a = 101.9 Å, b = 247.99 Å, c = 255.44 Å. Better crystals need to be obtained from further optimizations in the future.

Purification and crystallization of RIG-I HD-CTD in complex with a 14-bp dsRNA

While we were still working on the optimization of RIG-I HD-CTD complex crystals, recently, two groups have independently determined the crystal structures of RIG-I HD-CTD in complex with a 10- and a 14-bp dsRNA without 5'-ppp (110,111). Their structures provided insights into the roles of the helicase domain in RNA binding and RIG-I conformational change. Based on their methods, we generated a similar construct of RIG-I HD-CTD as a 6×His-SUMO-tagged protein which can be expressed in bacterial system, and successfully reproduced the complex crystal of RIG-I HD-CTD bound to the 14-bp dsRNA in 0.2 M NaSCN, 20% PEG 3350, and 3% trifluoroethanol.

Because the 10- and 14-bp dsRNA used by these two groups both lack the 5'-ppp preferred for signaling, it would still be interesting to see a structure of the HD-CTD in complex with a 5'-ppp dsRNA. Therefore we generated a 14-bp 5'-ppp dsRNA with GC-rich sequence by *in vitro* transcription using T7 RNA polymerase, and use it for cocrystallization. *E. coli* expressed RIG-I HD-CTD was purified by nickel affinity purification followed by complexes formation, SUMO-protease treatment, and gel filtration chromatography purification, similar to the steps that were used to process the HD-CTD/24bp dsRNA complex. Crystallization trials were set up using conditions around the one described above for the 14-bp non-ppp dsRNA. Hexagonal rod-like

crystals appeared in 0.2 M NaSCN, 17-19% PEG 3350, and 6% trifluoroethanol. Further optimizations of the crystals are needed as a future work.

EGCG is an inhibitor of RIG-I sinaling and ATPase activities without competing with RNA or ATP binding. Therefore it would be interesting to see where it binds to RIG-I and how it affects RIG-I structural conformation. We crystallize the RIG-I HD-CTD/14bp dsRNA complex in presence of different amount of EGCG with or without ATP, in conditions containing 0.2 M NaSCN, 15-25% PEG 3350, and 6% trifluoroethanol. The samples crystallized in similar shape as the crystals without inhibitor, but appeared to be thinner and longer. Further optimizations of the crystals are needed in the future work.

Discussion

In this study, we successfully developed an approach to purify the RIG-I HD-CTD and dsRNA complexes to high quality, and obtained their crystals, although we did not manage to collect diffraction data which is good enough for structure determination.

Recently, four groups have made breakthroughs in the studies of RIG-I activation mechanisms. Two groups determined the crystal structures of human RIG-I HD-CTD (RIG-I ΔCARDs) complexes, one with a 10-bp dsRNA and the other with a 14-bp dsRNA, at 2.5 Å and 2.9 Å resolutions, respectively (110,111). Another group presented the crystal structure of mouse RIG-I helicase domain (SF2 ATPase domain) in complex with the nucleotide analogue AMP-PNP, at 2.2 Å resolution (112). A fourth group reported the crystal structures of duck RIG-I constructs including the helicase domain

alone (3 Å) and in ternary complex with a non-ppp 19-bp dsRNA and an ATP analog AMP-PNP (4.2 Å) or ADP-AlF_3 (3.7 Å), the CARDs-helicase (ΔCTD, 3.4 Å), and full-length duck RIG-I (3.7 Å) (113). These recently published structures provide detailed information for understanding the mechanism of RIG-I activation by viral RNA.

Although using different fragments and origins of RIG-I, all four groups determined the crystal structures of the helicase domain, which contains two RecA-like domains (HEL1 and HEL2) together with a unique insertion domain in HEL2 (HEL2i) (Figure 5.3A). Intriguingly, there is a V-shaped structure which is composed of two long α helices, named the "pincer motif" (110) or "bridge" (113), that extends from HEL2, folds back to interact with HEL1, and then connects to the CTD (Figure 5.3B).

RIG-I completely surrounds the dsRNA molecule and encloses it in a ring structure, covering a region of 9–10 bp along the dsRNA (Figure 5.3). HEL1 faces the minor groove of the dsRNA and interacts with the RNA backbone of both strands, and the HEL2i domain contributes significantly to RNA binding. However, in the structure of RIG-I HD-CTD in complex with 10-bp dsRNA, the HEL2 domain is not in close contact with the RNA(Figure 5.3) (110), whereas in the structure of HD-CTD in complex with 14-bp dsRNA in the presence of an ATP analog ADP-BeF_3, the HEL2 domain directly interacts with the dsRNA (111), presumably because of the ATP-dependent conformational change coordinated by the pincer motif which functions like "the camshaft in an engine" (110).

Figure 5.3 The RIG-I HD-CTD (ΔCARDs) dsRNA complex (110)*.

(A) Schematic representation of the RIG-I protein. The RIG-I HD-CTD (ΔCARDs) construct is boxed.

(B) Overall structure of the RIG-I (ΔCARDs) dsRNA complex. Starting from the N terminus: HEL1 is in green, the bridge between the two helicase domains gray, HEL2 blue, the insertion domain (HEL2i) cyan, the pincer region (P) red, and the C-terminal domain (CTD) orange. The top strand of the RNA duplex (dsRNA10) is in magenta and the bottom strand pink.

(C) Side view showing the RNA interface with the central groove of RIG-I (ΔCARDs). This is a 90° rotation along the y axis relative to the orientation shown in (B).

(D) Solvent-accessible electrostatic surface views of RIG-I (ΔCARDs), shown with ± 10 k_bT/e_c.

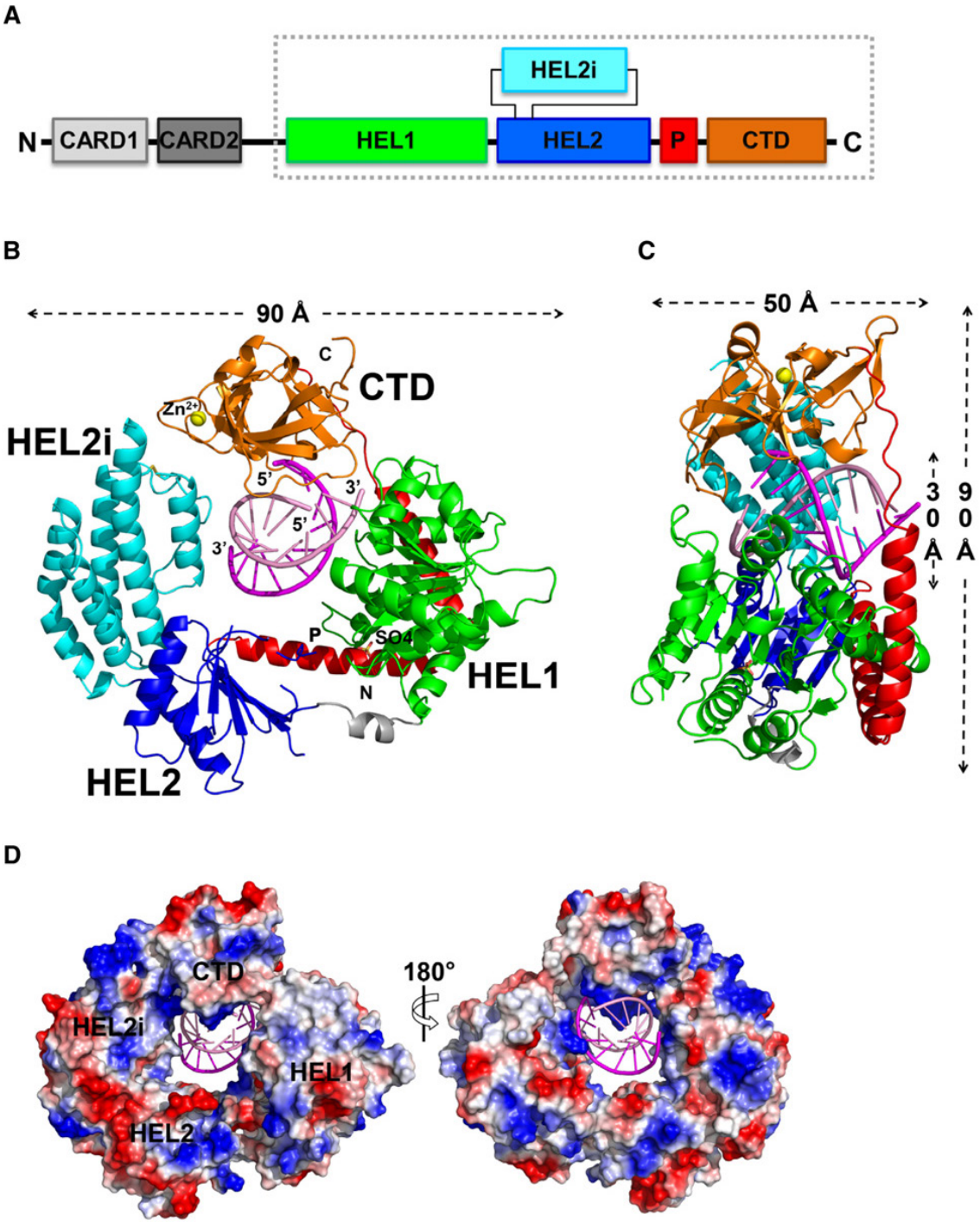
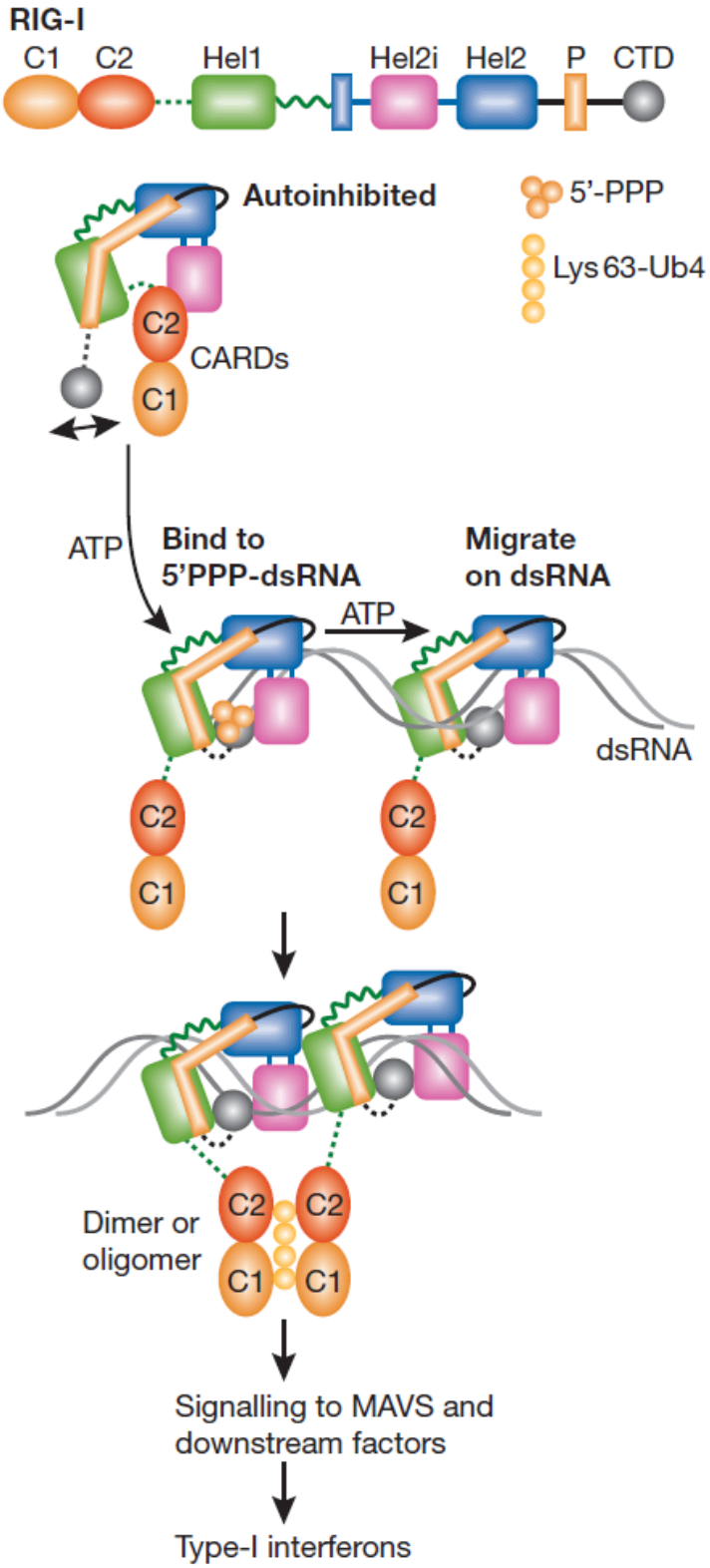

A
HEL2i
N
CARD1
CARD2
HEL1
HEL2
P
CTD
C
B
90 Å
C
CTD
Zn2+
HEL2i
5'
3'
5'
3'
P
SO4
HEL1
N
HEL2
C
50 Å
30 Å
90 Å
D
CTD
HEL2i
HEL1
HEL2
180°

Figure 5.4 Structural model of dsRNA-mediated activation of RIG-I (114)*.
RIG-I is an intracellular nanomachine that recognizes viral RNAs and elicits an innate immune response. In the absence of viral infection, RIG-I is autoinhibited, and the access of the CARDs, which are important for signal transduction, is limited sterically by the neighbouring helicase insertion domain (Hel2i) and the C2–Hel1 linker. The CTD recognizes 5'-triphosphate in dsRNA and initiates the binding of the helicase domains to the first turn of the double helix, where an ATP-binding site is assembled. ATP hydrolysis is required for the migration of RIG-I along the duplex RNA. The liberated CARDs bind Lys63 polyubiquitin chains and stimulate the massive aggregation of the mitochondrial protein MAVS, which ultimately leads to the induction of type I interferons and other antiviral molecules. C1/C2, CARD1/CARD2, caspase activation and recruitment domain 1/2; CTD, carboxy-terminal domain; dsRNA, double-stranded RNA; Hel1/Hel2, helicase 1/helicase 2; P, pincer domain.

*Reprinted with permission from "Structural insights into the activation of RIG-I, a nanosensor for viral RNAs" by Q.-X. Jiang, and Z.J. Cheng. 2012. *EMBO reports*, 13: 7-8. Copyright (2012) by Nature Publishing Group.


RIG-I
C1
C2
Hel1
Hel2i
Hel2
P
CTD
Autoinhibited
5'-PPP
Lys 63-Ub4
C2
CARDs
C1
ATP
Bind to
5'PPP-dsRNA
Migrate
on dsRNA
ATP
dsRNA
C2
C1
C2
C1
C2
C2
Dimer or
oligomer
C1
C1
Signalling to MAVS and
downstream factors
Type-I interferons

In addition to these ligand-bound structures, the structure of full-length RIG-I from duck in ligand-free state was also obtained (113). Comparison of these structures reveals a clear conformational rearrangement induced by RNA binding.

Based on these achievements, a revised model of RIG-I activation has been proposed (Figure 5.4). In the absence of RNA, full-length RIG-I adopts an autorepressed conformation in which the two CARDs interact with the HEL2i, preventing CARDs from polyubiquitination by TRIM25 that is critical for RIG-I activation. The autoinhibited RIG-I is open and flexible for the helicase domain and the CTD which was disordered in the crystal structure of full-length duck RIG-I in ligand-free state (113). The CTD is presumably available for viral RNA detection and the initial capture of 5'-ppp dsRNA would then recruit the RNA to the helicase. Subsequent cooperative binding of dsRNA and ATP causes a conformational rearrangement which is coordinated by the pincer motif, bringing the three helicase subdomains into close contact and creating a high-affinity complex between the helicase domain and the dsRNA. This dramatic change in conformation results in the releasing of the CARDs from their autorepressed state, allowing for modifications and interaction with IPS-1 (110,113). RIG-I may translocate along the RNA in an ATP-dependent manner (115) and RNA-induced dimerization or oligomerization of RIG-I may also be involved in the activation process (88).

Although those studies discussed above already made important discovers in understanding the mechanism of RIG-I activation, we have to notice that the dsRNA ligands they used in cocrystallization with RIG-I were all short dsRNA lacking the 5'-

ppp group, which may have low capacity for the induction of type I IFNs. Therefore it would still be worthwhile to continue our work in this Chapter. The crystals described in this study would provide more information if their structure can be determined. First, the 24-bp dsRNA we use to cocrystallize with RIG-I HD-CTD is longer than all those the other researchers used. It is twice the length of RIG-I footprint on dsRNA which is 9-10 bp and RIG-I forms a 2:1 complex with one protein binds to each end of the RNA. This may help us to understand the RNA-induced dimerization of RIG-I. Second, the 14-bp GC-rich dsRNA we use for HD-CTD cocrystallization contains the 5'-ppp group, which is crucial for RIG-I signaling induction. If structure of this complex can be obtained, we can compare it with the crystal structure of the non-ppp dsRNA complex to find out the differences between the two binding interactions and two RIG-I conformations. Furthermore, in our study, we use ATP instead of analogs, which may result in a different RIG-I conformational change because of the energy from ATP hydrolysis.

CHAPTER VI

CONCLUSIONS AND FUTURE DIRECTIONS

A better understanding of the function and activity of RIG-I as a critical antiviral innate immune receptor requires knowledge of the mechanisms associated with viral RNA sensing and signaling. There are three main questions that we would like to address in this study: 1) What are the structural features of viral RNA recognized by RIG-I? 2) How does RIG-I recognize viral RNA? 3) How does RNA binding activate RIG-I? To answer these questions, we applied both biochemical approaches and X-ray crystallography to analyze the interactions between RIG-I and various RNA. The results are summarized as follows.

To explore the RNA ligand specificity of RIG-I receptor, we performed biochemical binding studies on RIG-I CTD which is the domain that directly interacts with viral RNA. Gel filtration chromatography and EMSA experiments showed that RIG-I CTD binds to dsRNA with and without 5'-ppp group, as well as 5'-ppp ssRNA. Duplex RNA with both GC-rich and AU-rich sequences bind to RIG-I CTD very well, indicating that RIG-I recognizes RNA in a sequence independent manner. RIG-I CTD also binds to 5'- or 3'-overhanging dsRNA with 5'-ppp. However, SAP treatment of the overhanging dsRNA which removes the 5'-ppp dramatically reduced their binding with RIG-I CTD. Consistent with this, chemically synthesized dsRNA with 5'-overhangs failed to bind RIG-I CTD, while dsRNA with 3'-overhangs binds to RIG-I CTD with greatly reduced affinity. These demonstrated that the 5'-ppp group increases the affinity

of the dsRNA for RIG-I. The fact that RIG-I CTD does not bind to dsDNA also confirms that RIG-I is a viral RNA receptor.

Based on the elution profiles of gel filtration chromatography, we estimated that RIG-I CTD forms 2:1 complexes with blunt-ended dsRNA with or without 5'-ppp. AUC sedimentation velocity and equilibrium analyses confirmed the stoichiometry of RIG-I CTD binding to 5'-ppp dsRNA to be 2:1. In addition, a titration study by gel filtration chromatography analyzing the non-ppp dsRNA with increasing amount of RIG-I CTD from 1:1 to 4:1 also showed that the saturating ratio is 2:1.

Equilibrium and kinetic binding studies by SPR were conducted to determine the affinity and kinetic properties of RIG-I binding to 5'-ppp dsRNA, 5'-ppp ssRNA, and dsRNA without 5'-ppp. All three kinds of RNA bind to RIG-I CTD with high affinities. Among them, 5'-ppp dsRNA exhibits the highest binding affinity at 0.32 ± 0.05 nM for RIG-I and shows a slow association rate and a slow dissociation rate, indicating the requirement of conformational change in the interaction. In contrast, the binding of the non-ppp dsRNA shows faster association and dissociation rates.

The SPR results suggest that 5'-ppp dsRNA should be a more potent stimulator of RIG-I, and this has been verified by IFN-β luciferase reporter gene assays. The three types of RNA tested, non-ppp dsRNA, 5'-ppp dsRNA and ssRNA, are all able to stimulate RIG-I signaling activities in HEK 293T cells, and the 5'-ppp dsRNA has the highest activity when used at low concentration.

X-ray crystallography was used to obtain structural information in detail for understanding the binding interactions between RIG-I and dsRNA. We have crystallized

the three dsRNA complexes of human RIG-I CTD, two with 5'-ppp dsRNA and one with non-ppp dsRNA. In all structures, each dsRNA adopts a standard A-form double helical structure and recruits two CTD molecules on each terminal symmetrically. For 5'-ppp dsRNA binding, the contacts are made primarily through electrostatic interactions with a few nucleotides at the 5'-end and especially the 5'-ppp group which extensively interacts with the CTD at a positively charged cleft. On the other hand, the dsRNA without 5'-ppp interacts with the CTD through the first few base-paring nucleotides at both 5'- and 5'-ends. Comparison of the structures of RIG-I CTD bound to these dsRNA revealed that the orientation of two types of RNA relative to the protein is dramatically different. The dsRNA without 5'-ppp lacks the extensive interactions between the 5'-ppp moiety and the positively charged cleft, and it has to compensate by swinging towards the protein by about 15° to increase its contacts with the protein. Structural-based mutagenesis experiments revealed distinct but overlapping amino acids on the binding surface are involved in substrate recognition, and validated the importance of key residues in RNA recognition and RIG-I signaling. These studies demonstrated that RIG-I CTD is a versatile RNA-binding module using its positively charged binding surface to recognize different forms of RNA ligands. The diversity of virus types that can be detected by RIG-I may be explained by this versatility of substrate selection and recognition.

Comparing the crystal structures of RIG-I and LGP2 CTD complexes suggests that the two proteins use similar binding surfaces to interact with RNA, although the RNA is shifted in orientation. The dsRNA in LGP2 complex rotates along its helical axis and swings towards the CTD by about 15° relative to the position of the non-ppp dsRNA

in the RIG-I CTD complex. A similar CTD binding surface is also likely to be involved in MDA5 interaction with dsRNA. With no luck in crystallizing MDA5 CTD complex with dsRNA, we examine this prediction by a NMR titration study. The residues with perturbed chemical shift resonances were picked out as being involved in the interaction with RNA, and mapped onto the surface of MDA5 CTD crystal structure. The results showed a similar binding surface as RIG-I and LGP2 CTDs with positively charged residues. However, we notice that the critical residues involved in 5'-ppp moiety binding of RIG-I CTD are lacking in MDA5 and LGP2, suggesting a structural basis for the differences in RLR ligand specificity.

In the last part of this work, we also attempted to analyze the role of RIG-I helicase domain in RNA interaction and RIG-I activation by structural method. We expressed the human RIG-I HD-CTD in both baculovirus-insect cells system, and bacteria and purified the protein in complex with dsRNA. We crystallized HD-CTD bound to a 24-bp dsRNA without 5'-ppp as a 2:1 complex, in the presence of ATP. In addition, we also crystallized HD-CTD in complex with a 14-bp 5'-ppp dsRNA in the presence of either ATP or a RIG-I inhibitor EGCG. Unfortunately, because all these crystals were small and diffracted weakly, their crystal structures could not be obtained. Nonetheless, there are several other groups have independently determined the crystal structures of the helicase or HD-CTD in complex with short non-ppp dsRNA (110-113), provide insights into the mechanisms of RNA-induced ATP-dependent conformational change and activation of RIG-I. However, it would still be informative if the structures of the crystals described in this work could be determined in the future, because

currently there is no available structure of a 5'-ppp dsRNA bound helicase, which might show a distinct conformation from that of a non-ppp dsRNA bound structure. Furthermore, we still lack a crystal structure of full-length RIG-I in complex with which we could compare with the structure of full-length protein alone (113) to confirm the current RIG-I activation model.

The studies in this dissertation have been focused on RIG-I, the prototypical protein of RLRs. Questions regarding the pathogen recognition and signaling of other RLRs still remain unresolved. Future investigations would seek to define the RNA ligand specificity of MDA5 and LGP2, using the effective biochemical approaches described in this study. The activation mechanism of MDA5 is also of interest because it does not seem to be regulated by the autoinhibition mode. On the other hand, the CARDs-lacking LGP2 does not have a direct signaling activity. The positive and negative regulation of MDA5 and RIG-I signaling, respectively, suggest multiple functions of LGP2 that are undetermined. The general role of LGP2 in RLR-mediated antiviral signaling, and how it regulates the RLRs, whether by competing with ligand RNA binding or by direct interaction with RIG-I and MDA5, would be another important direction for future research.

REFERENCES

1. Hoffmann, J.A., Kafatos, F.C., Janeway, C.A. and Ezekowitz, R.A. (1999) Phylogenetic perspectives in innate immunity. *Science*, **284**, 1313-1318.
2. Cooper, M.D. and Alder, M.N. (2006) The evolution of adaptive immune systems. *Cell*, **124**, 815-822.
3. Janeway, C.A., Jr. and Medzhitov, R. (2002) Innate immune recognition. *Annu. Rev. Immunol.*, **20**, 197-216.
4. Janeway, C.A., Jr., Travers, P., Walport, M. and Shlomchik, M., J. (2005) Immunobiology : the immune system in health and disease. 6th ed. Garland Science, New York.
5. Medzhitov, R. and Janeway, C.A., Jr. (1997) Innate immunity: impact on the adaptive immune response. *Curr. Opin. Immunol.*, **9**, 4-9.
6. Martinon, F., Mayor, A. and Tschopp, J. (2009) The inflammasomes: guardians of the body. *Annu. Rev. Immunol.*, **27**, 229-265.
7. Fraser, I.P., Koziel, H. and Ezekowitz, R.A. (1998) The serum mannose-binding protein and the macrophage mannose receptor are pattern recognition molecules that link innate and adaptive immunity. *Semin. Immunol.*, **10**, 363-372.
8. Areschoug, T. and Gordon, S. (2009) Scavenger receptors: role in innate immunity and microbial pathogenesis. *Cell. Microbiol.*, **11**, 1160-1169.

9. Schwalbe, R.A., Dahlback, B., Coe, J.E. and Nelsestuen, G.L. (1992) Pentraxin family of proteins interact specifically with phosphorylcholine and/or phosphorylethanolamine. *Biochemistry*, **31**, 4907-4915.

10. Akira, S., Uematsu, S. and Takeuchi, O. (2006) Pathogen recognition and innate immunity. *Cell*, **124**, 783-801.

11. Gay, N.J. and Gangloff, M. (2007) Structure and function of toll receptors and their Ligands. *Annu Rev Biochem*, **76**, 141-165.

12. Kawai, T. and Akira, S. (2007) TLR signaling. *Semin. Immunol*, **19**, 24-32.

13. O'Neill, L.A. and Bowie, A.G. (2007) The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat. Rev. Immunol.*, **7**, 353-364.

14. Jin, M.S. and Lee, J.O. (2008) Structures of the toll-like receptor family and its ligand complexes. *Immunity*, **29**, 182-191.

15. Park, H.H., Lo, Y.C., Lin, S.C., Wang, L., Yang, J.K. and Wu, H. (2007) The death domain superfamily in intracellular signaling of apoptosis and inflammation. *Annu. Rev. Immunol.*, **25**, 561-586.

16. Pestka, S., Krause, C.D. and Walter, M.R. (2004) Interferons, interferon-like cytokines, and their receptors. *Immunol. Rev.*, **202**, 8-32.

17. Theofilopoulos, A.N., Baccala, R., Beutler, B. and Kono, D.H. (2005) Type I interferons (alpha/beta) in immunity and autoimmunity. *Annu. Rev. Immunol.*, **23**, 307-336.

18. Stetson, D.B. and Medzhitov, R. (2006) Type I interferons in host defense. *Immunity*, **25**, 373-381.

19. Pichlmair, A. and Reis e Sousa, C. (2007) Innate recognition of viruses. *Immunity*, **27**, 370-383.

20. Takeuchi, O. and Akira, S. (2010) Pattern recognition receptors and inflammation. *Cell*, **140**, 805-820.

21. Honda, K., Takaoka, A. and Taniguchi, T. (2006) Type I interferon gene induction by the interferon regulatory factor family of transcription factors. *Immunity*, **25**, 349-360.

22. van Boxel-Dezaire, A.H., Rani, M.R. and Stark, G.R. (2006) Complex modulation of cell type-specific signaling in response to type I interferons. *Immunity*, **25**, 361-372.

23. Maniatis, T., Falvo, J.V., Kim, T.H., Kim, T.K., Lin, C.H., Parekh, B.S. and Wathelet, M.G. (1998) Structure and function of the interferon-beta enhanceosome. *Cold Spring Harbor Symp. Quant. Biol.*, **63**, 609-620.

24. Wathelet, M.G., Lin, C.H., Parekh, B.S., Ronco, L.V., Howley, P.M. and Maniatis, T. (1998) Virus infection induces the assembly of coordinately activated transcription factors on the IFN-beta enhancer in vivo. *Mol. Cell*, **1**, 507-518.

25. Carey, M. (1998) The enhanceosome and transcriptional synergy. *Cell*, **92**, 5-8.

26. Tamura, T., Yanai, H., Savitsky, D. and Taniguchi, T. (2008) The IRF family transcription factors in immunity and oncogenesis. *Annu. Rev. Immunol.*, **26**, 535-584.

27. Samuel, C.E. (2001) Antiviral actions of interferons. *Clin. Microbiol. Rev*, **14**, 778-809, table of contents.

28. Berlanga, J.J., Ventoso, I., Harding, H.P., Deng, J., Ron, D., Sonenberg, N., Carrasco, L. and de Haro, C. (2006) Antiviral effect of the mammalian translation initiation factor 2alpha kinase GCN2 against RNA viruses. *EMBO J.*, **25**, 1730-1740.

29. Player, M.R. and Torrence, P.F. (1998) The 2-5A system: Modulation of viral and cellular processes through acceleration of RNA degradation. *Pharmacol. Ther.*, **78**, 55-113.

30. Malathi, K., Dong, B., Gale, M., Jr. and Silverman, R.H. (2007) Small self-RNA generated by RNase L amplifies antiviral innate immunity. *Nature*, **448**, 816-819.

31. Bieniasz, P.D. (2004) Intrinsic immunity: a front-line defense against viral attack. *Nat. Immunol.*, **5**, 1109-1115.

32. Le Bon, A. and Tough, D.F. (2002) Links between innate and adaptive immunity via type I interferon. *Curr. Opin. Immunol.*, **14**, 432-436.

33. Tough, D.F. (2004) Type I interferon as a link between innate and adaptive immunity through dendritic cell stimulation. *Leuk. Lymphoma*, **45**, 257-264.

34. Palm, N.W. and Medzhitov, R. (2009) Pattern recognition receptors and control of adaptive immunity. *Immunol. Rev.*, **227**, 221-233.

35. Akira, S. and Takeda, K. (2004) Toll-like receptor signalling. *Nat. Rev. Immunol.*, **4**, 499-511.

36. Kawai, T., Sato, S., Ishii, K.J., Coban, C., Hemmi, H., Yamamoto, M., Terai, K., Matsuda, M., Inoue, J., Uematsu, S. *et al.* (2004) Interferon-alpha induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nat. Immunol.*, **5**, 1061-1068.

37. Edelmann, K.H., Richardson-Burns, S., Alexopoulou, L., Tyler, K.L., Flavell, R.A. and Oldstone, M.B.A. (2004) Does Toll-like receptor 3 play a biological role in virus infections? *Virology*, **322**, 231-238.

38. Yoneyama, M. and Fujita, T. (2009) RNA recognition and signal transduction by RIG-I-like receptors. *Immunol. Rev.*, **227**, 54-65.

39. Fields, B.N., Knipe, D.M. and Howley, P.M. (2007) Fields virology. 5th ed. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia.

40. Kato, H., Sato, S., Yoneyama, M., Yamamoto, M., Uematsu, S., Matsui, K., Tsujimura, T., Takeda, K., Fujita, T., Takeuchi, O. *et al.* (2005) Cell type-specific involvement of RIG-I in antiviral response. *Immunity*, **23**, 19-28.

41. Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S. and Fujita, T. (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.*, **5**, 730-737.

42. Tanner, N.K. and Linder, P. (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol. Cell*, **8**, 251-262.

43. Tijsterman, M., Ketting, R.F. and Plasterk, R.H. (2002) The genetics of RNA silencing. *Annu. Rev. Genet.*, **36**, 489-519.

44. Fuller-Pace, F.V. (2006) DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res.*, **34**, 4206-4215.

45. Sun, Y.W. (1997) RIG-I, a Human Homolog Gene of RNA Helicase, Is Induced by Retinoic Acid During the Differentiation of Acute Promyelocytic Leukemia Cell. *Thesis, Shanghai Second Medical Univ*.

46. Kang, D.C., Gopalkrishnan, R.V., Wu, Q.P., Jankowsky, E., Pyle, A.M. and Fisher, P.B. (2002) mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 637-642.

47. Kovacsovics, M., Martinon, F., Micheau, O., Bodmer, J.L., Hofmann, K. and Tschopp, J. (2002) Overexpression of Helicard, a CARD-containing helicase cleaved during apoptosis, accelerates DNA degradation. *Curr. Biol.*, **12**, 838-843.

48. Andrejeva, J., Childs, K.S., Young, D.F., Carlos, T.S., Stock, N., Goodbourn, S. and Randall, R.E. (2004) The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN-beta promoter. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 17264-17269.

49. Yoneyama, M. and Fujita, T. (2008) Structural mechanism of RNA recognition by the RIG-I-like receptors. *Immunity*, **29**, 178-181.

50. Zou, J., Chang, M., Nie, P. and Secombes, C.J. (2009) Origin and evolution of the RIG-I like RNA helicase gene family. *BMC Evol. Biol.*, **9**, 85.

51. Yoneyama, M. and Fujita, T. (2007) Function of RIG-I-like receptors in antiviral innate immunity. *J. Biol. Chem.*, **282**, 15315-15318.

52. Takahasi, K., Yoneyama, M., Nishihori, T., Hirai, R., Kumeta, H., Narita, R., Gale, M., Jr., Inagaki, F. and Fujita, T. (2008) Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Mol. Cell*, **29**, 428-440.

53. Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., Ishii, K.J. *et al.* (2006) Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature*, **441**, 101-105.

54. Gitlin, L., Barchet, W., Gilfillan, S., Cella, M., Beutler, B., Flavell, R.A., Diamond, M.S. and Colonna, M. (2006) Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 8459-8464.

55. Loo, Y.M., Fornek, J., Crochet, N., Bajwa, G., Perwitasari, O., Martinez-Sobrido, L., Akira, S., Gill, M.A., Garcia-Sastre, A., Katze, M.G. *et al.* (2008) Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *J. Virol.*, **82**, 335-345.

56. Cui, Y., Li, M., Walton, K.D., Sun, K., Hanover, J.A., Furth, P.A. and Hennighausen, L. (2001) The Stat3/5 locus encodes novel endoplasmic reticulum

and helicase-like proteins that are preferentially expressed in normal and neoplastic mammary tissue. *Genomics*, **78**, 129-134.

57. Rothenfusser, S., Goutagny, N., DiPerna, G., Gong, M., Monks, B.G., Schoenemeyer, A., Yamamoto, M., Akira, S. and Fitzgerald, K.A. (2005) The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. *J. Immunol.*, **175**, 5260-5268.

58. Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y.M., Gale, M., Akira, S. *et al.* (2005) Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J. Immunol.*, **175**, 2851-2858.

59. Saito, T., Hirai, R., Loo, Y.M., Owen, D., Johnson, C.L., Sinha, S.C., Akira, S., Fujita, T. and Gale, M., Jr. (2007) Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 582-587.

60. Pippig, D.A., Hellmuth, J.C., Cui, S., Kirchhofer, A., Lammens, K., Lammens, A., Schmidt, A., Rothenfusser, S. and Hopfner, K.P. (2009) The regulatory domain of the RIG-I family ATPase LGP2 senses double-stranded RNA. *Nucleic Acids Res.*, **37**, 2014-2025.

61. Hornung, V., Ellegast, J., Kim, S., Brzozka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K.K., Schlee, M. *et al.* (2006) 5'-Triphosphate RNA is the ligand for RIG-I. *Science*, **314**, 994-997.

62. Pichlmair, A., Schulz, O., Tan, C.P., Naslund, T.I., Liljestrom, P., Weber, F. and Reis e Sousa, C. (2006) RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science*, **314**, 997-1001.

63. Bowie, A.G. and Fitzgerald, K.A. (2007) RIG-I: tri-ing to discriminate between self and non-self RNA. *Trends Immunol.*, **28**, 147-150.

64. Schlee, M., Roth, A., Hornung, V., Hagmann, C.A., Wimmenauer, V., Barchet, W., Coch, C., Janke, M., Mihailovic, A., Wardle, G. *et al.* (2009) Recognition of 5' triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus. *Immunity*, **31**, 25-34.

65. Schmidt, A., Schwerd, T., Hamm, W., Hellmuth, J.C., Cui, S., Wenzel, M., Hoffmann, F.S., Michallet, M.C., Besch, R., Hopfner, K.P. *et al.* (2009) 5'-triphosphate RNA requires base-paired structures to activate antiviral signaling via RIG-I. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 12067-12072.

66. Marques, J.T., Devosse, T., Wang, D., Zamanian-Daryoush, M., Serbinowski, P., Hartmann, R., Fujita, T., Behlke, M.A. and Williams, B.R. (2006) A structural basis for discriminating between self and nonself double-stranded RNAs in mammalian cells. *Nat. Biotechnol.*, **24**, 559-565.

67. Rehwinkel, J., Tan, C.P., Goubau, D., Schulz, O., Pichlmair, A., Bier, K., Robb, N., Vreede, F., Barclay, W., Fodor, E. *et al.* (2010) RIG-I detects viral genomic RNA during negative-strand RNA virus infection. *Cell*, **140**, 397-408.

68. Alexopoulou, L., Holt, A.C., Medzhitov, R. and Flavell, R.A. (2001) Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature*, **413**, 732-738.

69. Grunberg-Manago, M., Oritz, P.J. and Ochoa, S. (1955) Enzymatic synthesis of nucleic acidlike polynucleotides. *Science*, **122**, 907-910.

70. Takeuchi, O. and Akira, S. (2008) MDA5/RIG-I and virus recognition. *Curr. Opin. Immunol.*, **20**, 17-22.

71. Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K., Hiiragi, A., Dermody, T.S., Fujita, T. and Akira, S. (2008) Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. *J. Exp. Med.*, **205**, 1601-1610.

72. Cui, S., Eisenacher, K., Kirchhofer, A., Brzozka, K., Lammens, A., Lammens, K., Fujita, T., Conzelmann, K.K., Krug, A. and Hopfner, K.P. (2008) The C-terminal regulatory domain is the RNA 5 '-triphosphate sensor of RIG-I. *Mol. Cell*, **29**, 169-179.

73. Takahasi, K., Kumeta, H., Tsuduki, N., Narita, R., Shigemoto, T., Hirai, R., Yoneyama, M., Horiuchi, M., Ogura, K., Fujita, T. *et al.* (2009) Solution structures of cytosolic RNA sensor MDA5 and LGP2 C-terminal domains: identification of the RNA recognition loop in RIG-I-like receptors. *J. Biol. Chem.*, **284**, 17465-17474.

74. Li, X., Ranjith-Kumar, C.T., Brooks, M.T., Dharmaiah, S., Herr, A.B., Kao, C. and Li, P. (2009) The RIG-I-like receptor LGP2 recognizes the termini of double-stranded RNA. *J. Biol. Chem.*, **284**, 13881-13891.

75. Kawai, T., Takahashi, K., Sato, S., Coban, C., Kumar, H., Kato, H., Ishii, K.J., Takeuchi, O. and Akira, S. (2005) IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat. Immunol.*, **6**, 981-988.

76. Seth, R.B., Sun, L.J., Ea, C.K. and Chen, Z.J.J. (2005) Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappa B and IRF3. *Cell*, **122**, 669-682.

77. Xu, L.G., Wang, Y.Y., Han, K.J., Li, L.Y., Zhai, Z.H. and Shu, H.B. (2005) VISA is an adapter protein required for virus-triggered IFN-beta signaling. *Mol. Cell*, **19**, 727-740.

78. Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R. and Tschopp, R. (2005) Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature*, **437**, 1167-1172.

79. Michallet, M.C., Meylan, E., Ermolaeva, M.A., Vazquez, J., Rebsamen, M., Curran, J., Poeck, H., Bscheider, M., Hartmann, G., Konig, M. *et al.* (2008) TRADD protein is an essential component of the RIG-like helicase antiviral pathway. *Immunity*, **28**, 651-661.

80. Kawai, T. and Akira, S. (2008) Toll-like receptor and RIG-I-like receptor signaling. *Ann. N. Y. Acad. Sci.*, **1143**, 1-20.

81. Fitzgerald, K.A., McWhirter, S.M., Faia, K.L., Rowe, D.C., Latz, E., Golenbock, D.T., Coyle, A.J., Liao, S.M. and Maniatis, T. (2003) IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway. *Nat. Immunol.*, **4**, 491-496.

82. Sharma, S., tenOever, B.R., Grandvaux, N., Zhou, G.P., Lin, R. and Hiscott, J. (2003) Triggering the interferon antiviral response through an IKK-related pathway. *Science*, **300**, 1148-1151.

83. Honda, K. and Taniguchi, T. (2006) Toll-like receptor signaling and IRF transcription factors. *IUBMB Life*, **58**, 290-295.

84. Takeuchi, O. and Akira, S. (2009) Innate immunity to virus infection. *Immunol. Rev.*, **227**, 75-86.

85. Takahashi, K., Kawai, T., Kumar, H., Sato, S., Yonehara, S. and Akira, S. (2006) Roles of caspase-8 and caspase-10 in innate immune responses to double-stranded RNA. *J. Immunol.*, **176**, 4520-4524.

86. Gack, M.U., Shin, Y.C., Joo, C.H., Urano, T., Liang, C., Sun, L., Takeuchi, O., Akira, S., Chen, Z., Inoue, S. *et al.* (2007) TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature*, **446**, 916-920.

87. Gack, M.U., Nistal-Villan, E., Inn, K.S., Garcia-Sastre, A. and Jung, J.U. (2010) Phosphorylation-mediated negative regulation of RIG-I antiviral activity. *J. Virol.*, **84**, 3220-3229.

88. Binder, M., Eberle, F., Seitz, S., Mucke, N., Huber, C.M., Kiani, N., Kaderali, L., Lohmann, V., Dalpke, A. and Bartenschlager, R. (2011) Molecular mechanism of

signal perception and integration by the innate immune sensor retinoic acid-inducible gene-I (RIG-I). *J. Biol. Chem.*, **286**, 27278-27287.

89. Bamming, D. and Horvath, C.M. (2009) Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. *J. Biol. Chem.*, **284**, 9700-9712.

90. Schlee, M., Hartmann, E., Coch, C., Wimmenauer, V., Janke, M., Barchet, W. and Hartmann, G. (2009) Approaching the RNA ligand for RIG-I? *Immunol. Rev.*, **227**, 66-74.

91. Myszka, D.G., Wood, S.J. and Biere, A.L. (1999) Analysis of fibril elongation using surface plasmon resonance biosensors. *Methods Enzymol.*, **309**, 386-402.

92. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307-326.

93. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D: Biol. Crystallogr.*, **54**, 905-921.

94. Bailey, S. (1994) The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallogr. Sect. D: Biol. Crystallogr.*, **50**, 760-763.

95. Jones, T.A. and Kjeldgaard, M. (1997) Electron-density map interpretation. *Methods Enzymol.*, **277**, 173-208.

96. Lawrence, M.C. and Colman, P.M. (1993) Shape complementarity at protein/protein interfaces. *J. Mol. Biol.*, **234**, 946-950.

97. Saito, T., Owen, D.M., Jiang, F., Marcotrigiano, J. and Gale, M., Jr. (2008) Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature*, **454**, 523-527.

98. Chiu, Y.H., Macmillan, J.B. and Chen, Z.J. (2009) RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell*, **138**, 576-591.

99. Murali, A., Li, X.J., Ranjith-Kumar, C.T., Bhardwaj, K., Holzenburg, A., Li, P.W. and Kao, C.C. (2008) Structure and function of LGP2, a DEX(D/H) helicase that regulates the innate immunity response. *J. Biol. Chem.*, **283**, 15825-15833.

100. Ranjith-Kumar, C.T., Murali, A., Dong, W., Srisathiyanarayanan, D., Vaughan, R., Ortiz-Alacantara, J., Bhardwaj, K., Li, X., Li, P. and Kao, C.C. (2009) Agonist and antagonist recognition by RIG-I, a cytoplasmic innate immunity receptor. *J. Biol. Chem.*, **284**, 1155-1165.

101. Schlee, M. and Hartmann, G. (2010) The chase for the RIG-I ligand--recent advances. *Mol. Ther.*, **18**, 1254-1262.

102. Wang, Y., Ludwig, J., Schuberth, C., Goldeck, M., Schlee, M., Li, H., Juranek, S., Sheng, G., Micura, R., Tuschl, T. *et al.* (2010) Structural and functional insights into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nat. Struct. Mol. Biol.*, **17**, 781-787.

103. Li, X., Lu, C., Stewart, M., Xu, H., Strong, R.K., Igumenova, T. and Li, P. (2009) Structural basis of double-stranded RNA recognition by the RIG-I like receptor MDA5. *Arch. Biochem. Biophys.*, **488**, 23-33.

104. Muhandiram, D.R. and Kay, L.E. (1994) Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *J. Magn. Reson. B*, **103**, 203-216.

105. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277-293.

106. Bodenhausen, G. and Ruben, D.J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, **69**, 185-189.

107. Cordin, O., Banroques, J., Tanner, N.K. and Linder, P. (2006) The DEAD-box protein family of RNA helicases. *Gene*, **367**, 17-37.

108. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.*, **1**, 945-951.

109. Ranjith-Kumar, C.T., Lai, Y., Sarisky, R.T. and Cheng Kao, C. (2010) Green tea catechin, epigallocatechin gallate, suppresses signaling by the dsRNA innate immune receptor RIG-I. *PloS One*, **5**, e12878.

110. Luo, D., Ding, S.C., Vela, A., Kohlway, A., Lindenbach, B.D. and Pyle, A.M. (2011) Structural insights into RNA recognition by RIG-I. *Cell*, **147**, 409-422.

111. Jiang, F., Ramanathan, A., Miller, M.T., Tang, G.Q., Gale, M., Jr., Patel, S.S. and Marcotrigiano, J. (2011) Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature*, **479**, 423-427.

112. Civril, F., Bennett, M., Moldt, M., Deimling, T., Witte, G., Schiesser, S., Carell, T. and Hopfner, K.P. (2011) The RIG-I ATPase domain structure reveals insights into ATP-dependent antiviral signalling. *EMBO Rep.*, **12**, 1127-1134.

113. Kowalinski, E., Lunardi, T., McCarthy, A.A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D. and Cusack, S. (2011) Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell*, **147**, 423-435.

114. Jiang, Q.-X. and Chen, Z.J. (2012) Structural insights into the activation of RIG-I, a nanosensor for viral RNAs. *EMBO Rep.*, **13**, 7-8.

115. Myong, S., Cui, S., Cornish, P.V., Kirchhofer, A., Gack, M.U., Jung, J.U., Hopfner, K.P. and Ha, T. (2009) Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. *Science*, **323**, 1070-1074.

VITA

Name: Cheng Lu

Address: Department of Biochemistry and Biophysics, 2128 TAMU,
College Station, TX 77843-2128, USA

Email Address: journeyal@gmail.com

Education: B.S., Biological Sciences, Nankai University, Tianjin, China, 2007
Ph.D., Biochemistry, Texas A&M University, 2012

Publications:

1. **C. Lu**, C.T. Ranjith-Kumar, L. Hao, C.C. Kao, P. Li. (2011). Crystal structures of RIG-I C-terminal domain bound to blunt-ended double-strand RNA without 5' triphosphate. *Nucl. Acids Res.* 39:1565-1575.

2. **C. Lu**, H. Xu, C.T. Ranjith-Kumar, M.T. Brooks, T.Y. Hou, F. Hu, A.B. Herr, R.K. Strong, C.C. Kao, P. Li. (2010). The structural basis of 5' triphosphate double-stranded RNA recognition by RIG-I C-terminal domain. *Structure.* 18:1032-1043.

3. X. Li, **C. Lu**, M. Stewart, H. Xu, R.K. Strong, T. Igumenova, P. Li. (2009). Structural basis of double-stranded RNA recognition by the RIG-I like receptor MDA5. *Arch. Biochem. Biophys.* 488:23-33.